TITLE: CAN WEB INDICATORS BE USED TO ESTIMATE THE CITATION IMPACT OF CONFERENCE PAPERS IN ENGINEERING?

A thesis submitted in partial fulfilment of the
requirements of the University of Wolverhampton
for the degree of Doctor of Philosophy

October 2018

Signature----------

Date ----------

Abstract:

Although citation counts are widely used to support research evaluation, they can only reflect academic impacts, whereas research can also be useful outside academia. There is therefore a need for alternative indicators and empirical studies to evaluate them. Whilst many previous studies have investigated alternative indicators for journal articles and books, this thesis explores the importance and suitability of four web indicators for conference papers. These are readership counts from the online reference manager Mendeley and citation counts from Google Patents, Wikipedia and Google Books. To help evaluate these indicators for conference papers, correlations with Scopus citations were evaluated for each alternative indicator and compared with corresponding correlations between alternative indicators and citation counts for journal articles. Four subject areas that value conferences were chosen for the analysis: Computer Science Applications; Computer Software Engineering; Building & Construction Engineering; and Industrial & Manufacturing Engineering.

There were moderate correlations between Mendeley readership counts and Scopus citation counts for both journal articles and conference papers in Computer Science Applications and Computer Software. For conference papers in Building & Construction Engineering and Industrial & Manufacturing Engineering, the correlations between Mendeley readers and citation counts are much lower than for journal articles. Thus, in fields where conferences are important, Mendeley readership counts are reasonable impact indicators for conference papers although they are better impact indicators for journal articles.

Google Patent citations had low positive correlations with citation counts for both conference papers and journal articles in Software Engineering and Computer Science Applications. There were negative correlations for both conference papers and journal articles in Industrial and Manufacturing Engineering. However, conference papers in Building and Construction Engineering attracted no Google Patent citations. This suggests that there are disciplinary differences but little overall value for Google Patent citations as impact indicators in engineering fields valuing conferences.

Wikipedia citations had correlations with Scopus citations that were statistically significantly positive only in Computer Science Applications, whereas the correlations were not statistically significantly different from zero in Building & Construction Engineering, Industrial & Manufacturing Engineering and Software Engineering. Conference papers were less likely to be cited in Wikipedia than journal articles were in all fields, although the difference was minor in Software Engineering. Thus, Wikipedia citations seem to have little value in engineering fields valuing conferences.

Google Books citations had positive significant correlations with Scopus-indexed citations for conference papers in all fields except Building & Construction Engineering, where the correlations were not statistically significantly different from zero. Google Books citations seemed to be most valuable impact indicators in Computer Science Applications and Software Engineering, where the correlations were moderate, than in Industrial & Manufacturing Engineering, where the correlations were low. This means that Google Book citations are valuable indicators for conference papers in engineering fields valuing conferences.

Although evidence from correlation tests alone is insufficient to judge the value of alternative indicators, the results suggest that Mendeley readers and Google Books citations may be useful for both journal articles and conference papers in engineering fields that value conferences, but not Wikipedia citations or Google Patent citations.

Table of Contents

List of Tables

List of Figures

List of Appendices

Publications from this Thesis

Journal article

- Aduku, K.J., Thelwall, M., & Kousha, K. (2016), Do Mendeley reader counts reflect the scholarly impact of conference papers? An investigation of Computer Science and Engineering fields. *Scientometrics*. 112(1), 573-581.

Conference papers

- Aduku, K.J., Thelwall, M., & Kousha, K. (2015), Do mendeley reader counts correlates with citation counts? An investigation of Computer Science Applications, Computer Software Engineering, Industrial & Manufacturing Engineering and Building & Construction Engineering. June 29-July 3, **2015** 15th International **Conference** of the International Society for Scientometrics and Informetrics **Istanbul**, Turkey (Doctoral forum Presentation. p.8).

- Aduku, K.J., Thelwall, M., & Kousha, K. (2016), Do Mendeley reader counts reflect the scholarly impact of conference papers? An investigation of Computer Science and Engineering fields. *Proceedings of the 21th International conference on science and technology indicators – Peripheries, frontiers and beyond*, Valencia (Spain). 14-16 September 2016. p.17.

- Aduku, K.J., Thelwall, M., & Kousha, K. (2017), Can conference papers have information value through Wikipedia? An investigation of four engineering fields. *International conference on science and technology indicators – Open indicators: innovation, participation and actor-based STI indicators*, Paris 2017. 6-8 September 2017. (abstract ALT2. p.5).

Internal Conference Presentations

- Aduku, K.J., Thelwall, M., & Kousha, K. (2016), Do Patents cite conference papers as often as journal articles in engineering? An investigation of four fields. *Annual Research Conference-ARC 2016. University of Wolverhampton.*

- Aduku, K.J., Thelwall, M., & Kousha, K. (2017), Do Google Books citations reflect the scholarly impact of conference papers? An investigation of four engineering fields**.** *Annual Research Conference-ARC 2017. University of Wolverhampton.*

Acknowledgements

I dedicate this thesis to my Creator, the Almighty God and to the memory of my late Dad, Chief Aduku Agbu-Tsokwa Ashikushi and my late Mum, Mrs. Jumma Agyen Aduku

Special thanks to my lovely wife Mrs. Polina Joseph, for proffering solutions and supports to all challenges, and making sure am comfortable with writing this thesis.

Thanks to Professor Mike Thelwall for being more than a supervisor; he encourages me and taught me many lessons beyond this thesis.

Thanks to my second supervisor, Dr. Kayvan Kousha for his advice and support with this project.

Special thanks to my lovely children (Amayikai, Pa'abu, Yohana, Amazhende, Fojima and Chunma) for their resilience and perseverance in coping with the ever-challenging weather of United Kingdom. May God Almighty bless you all.

Thanks to the Nigerian government, Tertiary Education Trust Fund (Tetfund), for their financial support.

I would like to thank all the members of the Statistical Cybermetrics Research Group (SCRG) University of Wolverhampton, for their positive contributions to my research area and wonderful performance in the world of research.

Chapter 1: Introduction

1.1 Background

There have long been citation analyses of journal articles, books and monograph publications to assess their scholarly impact. In recent times, research evaluations have increased dramatically due to the involvement of universities, governments and funders seeking to assess the value of academic research processes (Wilsdon, 2016). Organisations need the knowledge from the values of academic research to: (a) help government and funding bodies to budget and spend their resources; (b) advise the universities on how to distribute their budgets within departments; (c) help tertiary institutions with staff appointments and promotions; and (d) help librarians decide whether to renew journal subscriptions. Whilst scholarly impact is traditionally assessed by peer judgements (normally the best source of evidence), this is sometimes supported by citation counts (Wouters & Costas, 2012).

Citation counts have been used in research evaluation in the belief that useful research is likely to be cited by future papers that exploit its ideas, methods or discoveries. Thus, other factors being equal, the more cited an article is, the more impact it has had. This is a substantial oversimplification, however, since work can be cited to criticise it and uncited work may have valuable societal impacts, such as curing diseases or triggering industrial innovations. Nevertheless, if used with care in support of qualitative judgements, citation counts can and do provide useful information to researchers and research managers. Recent years have seen increased dissatisfaction with citation counts as the major source of quantitative impact information, however, because of the rise of alternatives and an increasing recognition of the importance of non-scholarly impacts.

1.2 Research problem

Although scholarly impact is an important component of evaluations for the above purposes, research evaluation needs to consider non-scholarly impacts because these can be important research outcomes (Wilsdon, 2016). Citation analysis is the main quantitative method for helping to assess scholarly impact but does not directly assess non-scholarly impacts, such as commercialisation and societal value. Thus, indicators of alternative types of impact are needed to identify the impacts of research beyond the advancement of knowledge within science, such as for applications in engineering, economics, social policy, arts, culture, and the environment (Bornmann, 2012; Thelwall, 2012). These wider impacts of research publications may be reflected in new indicators that are derived from the web rather than from bibliometric databases (Cronin & Sugimoto, 2014; Priem, Taraborelli, Groth, & Neylon, 2010). Promising web indicators for non-standard research impacts include Google Patents citations, Wikipedia citations and Google Books citations. Many social media indicators have also been proposed, such as Twitter and Facebook citations.

A second limitation of citation counts is that they are slow to mount up, delaying the point at which it is feasible to conduct research evaluations. This issue can be partly resolved by alternative indicators, such as Mendeley readership counts (Gunn, 2013; Haustein & Siebenlist, 2011; Maflahi & Thelwall, 2016), that are much quicker to accumulate.

A third limitation of citation counts is that most prior studies have found them to be more useful for journal articles than for other types of scholarly outputs, such as conference papers and books. This is a problem in fields where conference papers or books are important or more important than journal articles. For example, conference papers are more cited than journal articles in some fields, underlining their importance (Goodrum, Mc Cain, Lawrence & Giles, 2001).

1.3 Research Aim and Objectives

The above three limitations of citation counts are combined in this thesis by assessing, for the first time, the value of four alternative indicators for conference papers in fields for which they are important. It investigates:

1. Fast academic impact with Mendeley readership counts.

Prior research has shown that Mendeley reader counts reflect scholarly impact for journal articles, and are probably supported by more evidence than any other altmetric (see: Maflahi & Thelwall, 2016; Thelwall, Haustein, Larivière & Sugimoto, 2013). Nevertheless, no previous study has investigated the value of Mendeley readership counts as academic impact indicators for conference-based fields.

2. Innovation and technological impact with Google Patents citation counts.

Citations from Patents to academic research can be useful indicators for innovation and technological impacts in some academic fields (Huang, Huang & Chen, 2014; Liaw, Chan, Fan & Chiang, 2014), and Google Patents is a new free source of this information (Wang, Zhang, & Xu, 2011). Google Patents has been shown to be a useful source of patents citations to journal articles in some fields (Kousha & Thelwall, 2017), but it is not known whether there are many patent citations to conference papers. This study fills this gap by investigating whether patents cite conference papers as often as journal articles in four engineering fields.

3. Informational and encyclopedia impact with Wikipedia citation counts.

Citations from Wikipedia to journal articles are rare but may indicate the general non-scientific impact of research (see: Brazzeal, 2011; Kousha & Thelwall, 2016). No prior studies have investigated the value of Wikipedia citations to conference papers and so it is not known whether such citations are common. This thesis fills this gap by investigating Wikipedia citations to conference papers in four engineering fields.

4. Book-based impact with Google Books citation counts.

Recent studies have shown that Google Books (GB) could be an important source of citations from books to journal articles and other books (Kousha & Thelwall, 2015a). although there are no studies that have investigated the same for citations to conference papers from Google Books.

The four objectives of this thesis are to address the above gaps by investigating Mendeley readership counts, Google Patent citations, Wikipedia citations, and Google Books citations

to conference papers in fields that value conferences. The objectives are addressed primarily by comparing the alternative indicators to Scopus citation counts using correlation tests. Correlations are useful as evidence that alternative indicators relate in some way to scholarly processes even though for three of the indicators (Google Patent citations, Wikipedia citations, and Google Books citations) their potential value lies in their ability to provide evidence of non-scholarly impact. This analysis is supported by qualitative explorations of outlier articles and papers that have attracted relatively high or low alternative indicator scores compared to citation counts. This qualitative component is designed both to investigate non-scholarly impacts and to test the robustness of the alternative indicators.

1.4 Terms and acronyms used in this thesis

The terms and acronyms below are defined for reference throughout the thesis.

Acronyms:

- API (Application Programming Interface): A software facility used by Scopus, Google Books and Mendeley that allows programs, such as Webometric Analyst, to automatically harvest data.
- DOI (Digital Object Identifier): A string of numbers, letters and symbols used to permanently identify an article or document.
- JIF (Journal Impact Factor): An indicator of journal impact. It was invented and defined by Eugene Garfield as the average number of times articles from a journal published in the past two years have been cited by journals indexed by the Clarivate Journal Citation Reports.
- PDF (Portable Document Format): A file format used to present and exchange documents.
- URL **(**Uniform Resource Locator): A reference to a web resource that specifies its location on a computer network and a mechanism for retrieving it. It is a type of Uniform Resource Identifier (URI).

Alternative indicators:

- Altmetrics: The original narrow definition of this term was for indicators of interest in academic documents derived from mentions in the social web and gathered through an API. A current wider informal meaning of this term covers all indicators of interest in academic documents derived from web data, irrespective of whether it is from the social web or gathered with an API. This thesis uses ‘web indicator’ instead for the latter meaning. Thus, in this thesis, the terms ‘social media metrics’ and ‘altmetrics’ refer to indicators extracted from social media tools for research assessment.
- Scopus**:** A citation database of peer-reviewed literature containing records of articles from scientific journals, books and conference proceedings.
- Mendeley: A free academic reference manager that can collect references, organise citations, and create bibliographies. It reports the number of users that have registered a document in their personal libraries.
- Mendeley reader/bookmark: A user of Mendeley that has added a given document to their personal Mendeley library, irrespective of whether they had read it first.

- Web indicator: A quantitative indicator of impact or interest in academic outputs derived from web data. The four web indicators used are: Mendeley readers, Google Patents citations, Wikipedia citations, and Google Books citations.
- Webometric Analyst: A free program (http://lexiurl.wlv.ac.uk) that can perform automatic searches with the Bing API, construct web searches for Wikipedia, patent and syllabus mentions, and can conduct and filter API searches through Mendeley, Google Books, and Scopus.

Statistics:

- Correlation: a bivariate statistic that measures the strength of association between two variables on a scale of -1 to +1. A value close to 1 indicates that the two variables tend to agree in rank or relative magnitude. A value close to -1 indicates that the two variables tend to disagree in rank or relative magnitude. A value close to 0 indicates that the two variables tend to neither agree nor disagree in rank or relative magnitude, as would be the case for two random variables.
- Spearman correlation: Spearman rank correlation is a non-parametric test that is used to measure the degree to which two variables are in the same rank order.
    - Weak correlation: A positive correlation of r= 0.001 to 0.999.
    - Low correlation: A positive correlation of r=0.1 to 0.299.
    - Moderate correlation: A positive correlation of r= 0.3 to 0.499.
    - Strong correlation: A positive correlation where values of r=0.5+ (Cohen, 1988).
- P values from correlation tests: From a frequentist perspective, these values reflect the probability of incorrectly rejecting the null hypothesis (H_0) that two samples are drawn from a population with a correlation of zero. A p-value of 0.05 therefore suggests that there is a less than 5 in 100 (5%) chance of wrongly concluding that there is a non-zero correlation between a citation count and another web indicators in the whole population, if the population correlation is in fact 0.
- Statistically significant: This thesis uses the 5% threshold of significance, with a frequentist perspective. Thus, a null hypothesis is rejected at the 5% level if only 5% of such hypothesis would be incorrectly rejected, if they were correct.
- Confidence interval: A confidence interval reflects how much uncertainty there is with any statistic. From a frequentist perspective, a 95% confidence interval from a sample would be expected to contain the population mean 95% of the time.
- Margin of error: The margin of error is the range of values below and above the sample statistics in a confidence interval.

Citation analysis and research evaluation:

- Scholarly impact: A contribution to academic knowledge, including in the form of methods, ideas and discoveries. Scholarly impact could typically be expected to translate into citations from other scholarly outputs, such as journal articles, conference papers and books.
- Knowledge transfer: Transmission of know-how (knowledge) which enables the recipient enterprise to manufacture a product or service or for any kind of organisation to improve their processes or outputs. Scholarly knowledge transfer

seeks to organize, create, capture or distribute knowledge and ensure its availability for future users (Schmoch, 1993).

- Citation count: The number of times that a scholarly document has been referenced in other scholarly documents, as recorded in a citation index.
- Subject category: A coherent collection of academic journals that have a similar topic focus, such as the Scopus subject categories Computer science applications, Software Engineering, Building & Construction Engineering and Industrial & Manufacturing Engineering. Other subject categories include that of the Web of Science and Science-Metrix.

1.5 Thesis Structure

This thesis comprises of seven chapters.

Chapter One gives an overview of the research, including the background of the study, the main research problems, the aims and objectives, and the key terms and acronyms used.

Chapter Two reviews literature related to this thesis, starting from citation-based indicators and the use of citation analysis in research evaluation. It also introduces web indicators, the impacts and challenges of using web indicators for scholarly and non-scholarly impacts. The chapter concludes by identifying gaps in the prior literature.

Chapter Three introduces the research questions addressed in this thesis.

Chapter Four forms the methodology, outlining the methods used in the four studies: Collecting Mendeley readership counts, Scopus citations, Google patents citations, Wikipedia citations, Google Books citations; and methods to evaluate the data collected using correlation tests, confidence intervals, hypothesis tests and outlier analyses.

Chapter Five gives the results of the analyses and Chapter Six evaluates them in the context of the prior literature.

Chapter Seven is the Conclusion and is divided into five parts. After a brief introduction, the findings for each research questioned are summarised. The contributions of the study to scientific community, especially Altmetrics and bibliometrics research are discussed. The next part discusses the implication of the findings and summary of the indicators that are acceptable for the study and finally, some recommendations are put forward for future studies.

Chapter 2: Literature Review

This chapter reviews literature related to both citation analysis and alternative indicators. As discussed above, although conventional citation databases are commonly used for monitoring the citation impact of published research, they may not reflect its wider impacts. For example, conference papers can be cited by publications which are not indexed by conventional citation databases (books, patents, encyclopaedia entries), and can be used for wider scholarly reasons which cannot be traced through citations, such as for educational applications, as well as for applications outside of academia.

Alternative web indictors were introduced in an attempt to address the limitations of citation counts. Early web impact indicators were mainly based on web links (Ingwersen, 1998), mentions of academics' names (Cronin *et al*., 1998) or web citations to scientific publications in online documents (Vaughan & Shaw, 2003). These can all identify impacts beyond conventional citation analysis (Cronin, 2001). Currently, web citations (citations from digitised scholarly documents), and social media metrics (known as altmetrics), have been proposed to assess the broader impact of research. This chapter reviews empirical evidence about web citation and altmetric indicators for academic publications. Although this thesis focuses on conference papers, most of the literature review deals with journal articles since there is little research about conference papers so far.

2.1 Citation analysis

The concept of academic citation analysis dates to 1927, when Gross and Gross (1927) first assessed the importance of scholarly work using citation counts. Since then, citation analyses has been used to support research assessment in many different ways, including: national science policies and disciplinary development in supporting decision makers in science and technology; for departments and research laboratories to aid authenticating research results; and for books and journals in supporting scholarly impact and individual scientists (Garfield, 1972; Nicolaisen, 2002).

Van Raan (2004) categorized researchers that are interested in measuring scientific impact into two groups. The first group assesses the value bibliometric analyses for research evaluation, such as by assessing how it correlates with scientific assessments, such as academic ranking, research grants, peer judgements, and awards; whilst the second group is interested in exploiting citation counts as an indicator of scholarly impact for an application.

2.1.1 Citation-based indicators

Although raw citation counts are sometimes used in research evaluation, they are also often combined into an indicator using a mathematical formula. Examples include the h-index, average citation counts and field normalised average citation counts.

The Journal Impact Factor (JIF) is a well-known citation-based indicator (Garfield, 1995 in Holmberg, 2015). The JIF is the ratio of citations to a journal divided by the number of articles published in that journal during the 2 years prior to the calculation of the JIF (Holmberg, 2015, p.24). The JIF has been popular because of its simple definition and its

publisher's (now Clarivate Analytics) good reputation (Bollen *et al*., 2009). Nevertheless, it has been widely criticised and is currently under attack by the San Francisco Declaration on Research Assessment (DORA), which is an initiative to prevent the inappropriate uses of JIF in research management. Boell and Wilson (2010) argued that the JIF has been inappropriately used for assessing the performance of researchers, institutions and research groups. Smith (2012) observed that the JIF has changed scientific publication strategies and is the most controversial bibliometric indicator.

Not all of the valuable dimensions of journals can be reflected by the JIF and so a good journal may have a low score (Dellavalle *et al*., 2007). Bollen *et al*. (2006) combined the JIF and weighted PageRank to present an indicator for the prestige of a journal, as they believed that the JIF reflected only the popularity of a journal rather than its prestige. Many other alternative journal impact indicators have also been proposed, including some using altmetric data. These are not the main focus of this thesis, however.

For individual academics, standard bibliometric indicators include the number of publications produced, the total number of citations to these publications and the average number of citations per publication. Hirsch (2005) and Egghe (2006) suggested the h-index and the g-index respectively, for evaluating research output at the level of the author in a way that combines their productivity (number of publications) with their scholarly impact (citations). These hybrid indicators have been widely criticised by the scientometric community and now seem to be in disfavour.

For departments, institutions and countries, field normalised citation impact indicators, such as the Mean Normalised Citation Score (MNCS) (Waltman *et al*., 2011) and the Mean Normalised Log-transformed Citation Score (MNLCS) (Thelwall, 2017a) reflect the average citation rate of the publications produced by each group, taking into account field and year differences in citation rates. The MNCS, and variants, are a standard tool for comparing the impacts of large groups of researchers.

A simple example of an application of citation analysis that includes the use of both counts and averages is a study by Ani (2017) of Library and Information Science (LIS) in universities in Nigeria. Field normalisation was not used because the cited papers were from the same broad area. The study ranked universities by total citations (University of Ibadan: 91; Delta State University: 38; University of Lagos: 26; Obafemi Awolowo University: 25; Federal University of Technology Akure: 21) and average citation counts (Federal University of Technology Akure: 4.2; University of Lagos: 2.9; Obafemi Awolowo University: 2.5; University of Calabar: 2.5).

2.1.2 Problems with using citation analysis for research evaluation

Evaluating research using citation analysis is controversial, with many opponents that consider them to be fundamentally flawed (MacRoberts & MacRoberts, 1996). MacRoberts and MacRoberts (2010) argued that publications that are not cited may still have had an impact and other types of research activities may also have had some impact on the scientific community or on the public in general. Citations can only reflect scholarly impact rather than

applied impact. Even proponents of citation analysis accept that they have many limitations for research assessment (Moed, 2005).

At the heart of citation analysis is a belief that citations tend to be positive indications that a study has influenced subsequent research. Nevertheless, the likelihood of being cited depends on many factors, such as the following.

1. **Time**. The continued increase of scientific publications causes citations to become more numerous from year to year. Van Raan (2004) argued that there are bound to be fewer citations to recent publications than to older ones in the longer term because of the expansion of science overall, although the most recent publications have had less time to be cited and will have the fewest citations.

2. **Field**. There are variations in citation practices between subject areas and even within different areas or sub areas (Ziman, 2000). Citation counts may also be related to the size of a field (Moed *et al*., 1985). Thus, articles in smaller fields may tend to attract fewer citations than articles in general fields.

3. **Journal properties**. Smart and Waldfogel (1996) showed that the number obtained can be influenced by the order in which articles are listed within a journal issue. For example, the first articles tend to receive more citations in a scientific journal than later ones.

4. **Article properties**. Longer articles tend to be more cited than shorter ones. MacRoberts and MacRoberts (1996) argued that there are different citation trajectories between methodology articles, review articles and research articles and between articles, chapters and books.

5. **Author/reader properties**. Van Raan (2005b) observed that culture and language barriers can affect the citations to papers. Sandstrom *et al*. (2005) showed that authors who are familiar with each other tend to cite one another's work.

6. **Availability**. Lawrence (2001) argued that open access publications are more likely to be more cited.

7. **Technical issues**. The accuracy of citation analyses depends on the raw material used. For example, 48% of the references in one medical journal were incorrect (Van Raan, 2005b). Many authors might not check their references before publication and some do not even read through the whole contents of the documents cited. Similarly, Eichorn and Yankauer (1987) found that 31% of 150 references had citation errors, with 10% being difficult to trace. Citation counts used in research evaluations presumably exclude these incorrect matches.

8. **Database coverage**. Citations from Web of Science and Scopus reflect their different coverage of scientific journals and neither Web of Science nor Scopus are good at covering the social sciences and humanities (Ardanuy, 2013). Google Scholar has better coverage but is difficult to use in research evaluations due to the lack of a method for

automatically harvesting citation counts on a large scale (with the minor exception of the Publish and Perish software).

Whilst some of the above factors can be taken into account in the design of indicators (e.g., time, field, article type), others are impractical or impossible to resolve (e.g., journal, some technical issue, author properties).

The above issues can have important implications for research evaluation. For example, Holmberg (2015) argued the time delay in citations can cause wider problems in research evaluation. For instance, when analysing research conducted at a specific university, some of the authors may not be at that university a few years later. When evaluating the work of a researcher, it is possible that their most important work has just been published and therefore has not yet received any recognition in the form of citations. Perhaps more fundamentally, rewarding research through citation-based decisions focuses on past success rather than current research or future potential.

2.2 Non-journal publication citations

Although citation counts are widely used to support research evaluation, they can only reflect academic impacts, whereas research can also be useful outside of academia. Thus, perhaps the most fundamental limitation of citation analysis is that citations cannot reflect important non-scholarly impacts of research. The altmetrics movement has argued for the need to adopt new types of indicators to supplement them, provide new perspectives for research evaluation (Priem et al., 2012; Priem & Hemminger, 2010; Torres-Salinas et al., 2013). This has led to the creation of new alternative indicators that are not based on citations from journals.

Alternative indicators derived from patents (Trajtenberg, 1990), the general web (Thelwall & Kousha, 2015), or the social web (Priem *et al*., 2010), have been proposed to reflect the different types of impacts that documents can have. Examples of web/social web indicators investigated in this thesis include Mendeley readership counts, Wikipedia citations, Google Patents citations and Google Books citations. The value of these alternative indicators has been assessed primarily through correlation tests on the basis that robust new indicators should have positive correlations with citation counts, even if they reflected different types of value (Sud & Thelwall, 2014).

The following subsection reviews the most complex source of alternative indicators, patents, and subsequent sections explores more general web indicators.

2.2.1 Patent citations

Patents are official documents granted to inventors to make use of and to exploit their inventions for a limited period within a country or company. Patents are commercial documents in the sense that their ownership has financial value for useful inventions. Patents can cite academic research and these citations can be used as evidence of the commercial usefulness of the cited papers. The term patent citations can refer to citations to patents from patents or other documents or citations from patents to patents or other documents. Despite being the main source of quantitative data on academic impact, scholarly citations are unable

to reflect the uptake of research within industry. In contrast, patent citations can reflect the commercialisation of research and provide information on the development of new technology and innovation on a global scale.

A patent citation index was proposed by Arthur H. Seidel in the 1940s in the Journal of the Patent Office Society. Although patent citation analysis in the past has been too labour intensive to be widely used (Verbeek *et al*., 2002), it has been a standard bibliometric technique for over 20 years in the USA (Narin, 1994) and is helped by the increasing availability of patent databases, including some that are free online.

According to Narin *et al*. (1984), patent citation analysis can identify relationships among patents and can be used to determine technological knowledge flows between industrial inventors. Academic knowledge transfer can be also reflected by citations from patents to scientific articles, monographs and conference proceedings (Tussen *et al*., 2000; Verbeek *et al*., 2002). This technical knowledge from patents citations can help with marketing strategy, licensing, technical forecasting and future information for detecting competitors. Citations from patents can also be used to assess the industrial relevance of academic research areas (Wang *et al*., 2011). Patent citation analysis has been used to identify innovative technology (Érdi *et al*., 2013), to evaluate research performance (Moed, 2009), and as a potential indicator of economic value (Lanjouw & Schankerman, 2004). More generally, patent citations have been used to investigate knowledge flows in relation to technology (Fleming *et al*., 2007; Sorenson *et al*., 2006). They can also be used to analyse the productivity of countries, assignees and inventors (Narin, 1994). Patent analysis can estimate the relationship between the technological advances in an organisation and the likelihood of future technological advances in a field of technology. Patent citations have been shown to be important for securing grants in some areas, as observed in the study of Kapoor *et al*. (2016) for European Wind Industry Patents.

Patent citations have been used within scientometrics to explore the relationship between innovations and the patent system (Murray, 2002), as business impact indicators (Moskovkin, Shigorina, & Popov, 2012), and to explore impact within engineering (White, 2015). Previous bibliometric studies using patent citations have found that manual searches of patent databases are time-consuming. These include the Derwent Patents Index and Google Patents, neither of which contains an academic citation index. Perhaps because of this, many patent citations studies are constrained to a small window of years and disciplines until the arrival of a semi-automatic method of extracting and filtering patent citations via the Bing API (Kousha & Thelwall, 2015c)

Patent citations have unique problems that are not present in citations from journal articles. Alcacer and Gittelmam (2006) found that 63% of patent citations were added by the patent examiners, while only a small proportion of the citations contained the inventor's knowledge (Criscuolo & Verspagen, 2008). This undermines the use of patent citations for evidence of direct knowledge flows. In response, Lopez (2010) argued that the purpose of patent citation analysis is to establish a technological relationship between countries, assignees, inventor and techniques.

A major practical disadvantage of early bibliometric studies of patent citations is that it takes a long time for manual searches of patent databases, such as Derwent Patents Index and Google Patents, neither of which contains an academic citation index (Verbeek *et al*., 2002; Shirable, 2014). Perhaps because of this, many patent citations studies are restricted to a small window of years and few disciplines (Callaert *et al*., 2006; Meyer *et al*., 2010). This process can be semi-automated for efficiency, however (Lopez, 2010; Callaert *et al*., 2012), using repeated downloads of all the patents or collections of all the patents as a first step.

Abbas *et al*. (2014) observed the need to develop automated tools for patent analysis, considering the accumulated growth of patents. Organizations use different tools for patent analysis, and most of these tools seem to have the capability to measure lots of tasks associated with the patent analysis. For example, most patent analysis tools can be used for forecasting technology trends, conducting strategic technology planning, detecting patent infringement, determining patents quality, identifying technological inventions and filling up any existing patent vacuums. In addition, analysing patents helps organizations to determine the originality of their inventions and to identify the Intellectual Property (IP) and strengths and weaknesses of other competitors. The author observed that IP information can be used in estimating the developments of an organization with patent inventions for a specific time interval.

2.3 Social media metrics: Overview

In 1992, the Web was made available for public information dissemination and private knowledge. Early simple web indicators were soon calculated, including the size of the web, and parallels between the web and academic papers have led to many new web indicators motivated by bibliometric indicators (Rousseau, 1999: in Katz, 2006). The rise of the social web then led to a new class of social media metrics.

In this dissertation, the terms ‘social media metrics’ and ‘altmetrics’ refer to indicators extracted from social media tools for research assessment, although some scholars have used other terms for the same concept such as ‘Scientometrics 2.0’. For example, Priem *et al*. (2010) published the Altmetrics Manifesto (http://altmetrics.org/manifesto/) which began by stating that; ‘No one can read everything’. They argued that altmetrics could reflect the broad, rapid impact of scholarship in the growing online environment. They called for more tools and research based on altmetrics. The Manifesto emphasizes the potential of altmetrics to help scholars to filter the increasing volume of papers so that they could more easily identify the most valuable research. This would also help with assessing the impact of research. The Manifesto also claimed that altmetrics could point to the broader impacts of research: ‘Altmetrics expand our view of what impact looks like, but also of what’s making the impact’.

Rousseau and Fred (2013) argue that although the idea of altmetrics is good, the name given to the concept was a bad one. The authors highlighted that what is an alternative today might not be an alternative in ten years. Rousseau and Fred (2013) proposed the term ‘influmetrics’ as a substitute for altmetrics, but this has not been widely adopted. In addition, the authors

argued that using social media for scholarly communication could bridge the gap between academia and everyday life much faster and easier than the traditional citation counts that are slow in nature and undergo delay in disseminating scholarly information.

Multiple definitions have been proposed for altmetrics. Priem (2014) defines altmetrics as 'the study and use of scholarly impact measures based on activity in online tools and environments'. Shema *et al*. (2014) define altmetrics as 'web-based metrics for the impact of scholarly material, with an emphasis on social media outlets as sources of data'. Adie and Roe (2013) stated that 'altmetrics presented an alternative to the current practice of relying only on citation counts and journal impact factors for the quantitative analysis of impact by introducing new complementary approaches and sources of data'. This emphasizes that altmetrics do not stand for alternative metrics, but an alternative view on research indicators and approaches to research evaluation in general. The greatest potential of altmetrics may be in complementing citation-based indicators and in providing information about previously hidden aspects of the underlying impact of scientific work. According to Liu and Adie (2013), altmetrics could reflect broader views of research impact such as: (1) a multidimensional view of scholarly communication in the networked digital world and (2) a different view of the impact of research beyond scientific communication.

While some altmetrics are created from types of scholarly communication and are therefore connected to research work, others are based on data that is created at least in part by the public and may indicate broader societal impact (Bornmann, 2012; Bornmann, 2014). Nevertheless, whilst some altmetrics have the potential to be used as scientific impact measures, others have not. Even though many different data indicators are referred to as altmetrics, they may reflect different types of impact, or none.

2.3.1 Scholarly use of the social web

A decade ago, researchers were not using social media extensively for scholarly communication. In a study from 2010. about how academics in the United Kingdom used social media, only 13% of the 1308 respondents used social media frequently, 45% occasionally did and 39% did not use social media at all (Procter *et al*., 2010). Several studies have shown that those who used social media used it for communicating their work, connecting with colleagues and learning what others were doing. For example, the investigation of Rowlands *et al*. (2011) on social media uptake of about 2000 respondents showed that the most popular social media tools for scholarly work were those used for collaborative authoring, conferencing and scheduling meetings. A survey from one university in Finland had similar results, with 40% of the respondents that used social media reporting that they were useful for scholarly communication (Gu, & Widén-Wulff, 2011). Ponte and Simon (2011) reported moderate or low adoption of social media tools, with 42% of the respondents using wikis, 39% using blogs, 35% using social networks, 26% using social bookmarking sites and 18% using microblogs, such as Twitter. Whilst the uptake of social media tools has been relatively low, those that use social media experience several benefits with them. Gruzd *et al*. (2012) reported that researchers use social media tools to keep up with developments in their area of research, and that academic social networking sites (e.g.

ResearchGate) are particularly well suited for discovering new information and keeping up with developments in the field.

An online presence is important for researchers and these presences may be exploited for new altmetrics. Bar-Ilan *et al*. (2012) investigated the web presence of 57 presenters at the STI conference in 2010. They discovered that 70% of the respondents had a profile on LinkedIn, 23% on Google Scholar, and 16% on Twitter. In a follow-up study of the respondents, over half stated that they used some social media tools professionally (Haustein *et al*., 2014). While most (68%) had a profile on LinkedIn, the academic social networking sites Academic.edu and ResearchGate, and the social bookmarking site Mendeley were each used by about a fifth of the respondents. Similarly, Mas-Bleda *et al*. (2014) investigated to what extent highly cited European researches have public profile pages on Google Scholar and web presence on Mendeley, Academia.edu, LinkedIn and SlideShare. LinkedIn was the most popular (about 6% on average). Only 28% of the investigated researchers used any of the investigated sites.

Altmetrics rely on real-time data and interactions that can be quantified and measured immediately (Galloway *et al*., 2013). Existing altmetrics have used a variety of data sources including article downloads (Bollen *et al*., 2008), views and saves, as well as tweets, blogs, bookmarking sites and wikis. These are all used by scholars to communicate different kinds of research impact (Cronin, 2013).

Zahedi et al. (2017), analysed the relationship between the characteristics of different document types and social media web-based indicators and the influence of citation counts for Mendeley readership counts, tweets, Facebook posts, mentions in blogs and mainstream media for 1,339,279 papers published in 2012 in Web of Science. The study revealed that 84.2% of these articles were saved in Mendeley, whilst 81.7% were cited in the Web of Science.

The social media coverage for other platforms was much lower, with only 22.6% of papers receiving at least one tweet, 5.2% being shared publicly on Facebook, 2.3% mentioned in blog posts, and 1.1% discussed by mainstream media. The, most commonly cited or saved documents on Mendeley were reviews and articles, while editorial material and news items were mostly on cited Twitter, Facebook, blogs, and mainstream media. The study also shows that both coverage and density were higher for reviews and articles, but editorials and news items were also frequently saved by Mendeley users. The correlations between citation counts and the various social media web-based indicators were also investigated for Mendeley readership counts (0.585), Twitter (0.279), Blogs (0.159), Facebook (0.142) and mainstream media (0.115). For document characteristics, Mendeley readership counts have a positive correlation (0.471) with the number of references made, the number of pages (0.287) and title length (0.080).

2.3.2 Advantages and disadvantages of altmetrics

Holmberg (2015) lists some potential benefits of using altmetrics for research evaluation as follows:

1. While citations take a long time to accumulate, social media metrics can be generated immediately after publication.
2. Altmetrics can capture the impact and visibility of a much wider range of research products.
3. Most altmetrics use openly available data and therefore they provide a level of transparency not possible when using citation-based indicators from proprietary sources. Although most of the data is open because of the discretion of the social media sites, this could change at any time.
4. Altmetrics may favour open access publishing, with open access papers gaining more visibility and impact online.

The author stated that a methodological problem with altmetrics comes from the very dynamic nature of the web in general and social media in particular. For instance, websites may disappear without a warning, while new sites enter the field. The author however highlights the following possible challenges to altmetrics:

1. The need for combining all the different versions of the same research output that may appear in many different online places and possibly with different identifiers, or even without any unique identifiers.
2. Altmetrics have no standards for both data and for practices for collecting, aggregating, and presenting data. Providers of altmetrics can collect data from many different sources and report the altmetrics in many different ways.
3. A theoretical problem with altmetrics arises from the lack of understanding of what altmetrics really mean and what, if anything, these new indicators can tell about research impact.
4. A technical problem comes from the difficulties in normalizing the data. Thus, many disciplinary differences in the publishing and citation traditions have been discovered, and some of these seem to apply in a social media context as well, but perhaps for different reasons.
5. Altmetrics could be criticized for being easy to manipulate. For instance, creating hundreds of automated Twitter accounts, but as is the case for citations, the large volume of social media data may allow filtering out of the manipulated outliers. The author claimed that altmetrics are not yet extensively manipulated, but that may change quickly if and when altmetrics are used in decision-making about, for instance, research funding, at which point the mechanisms to detect manipulation need to be in place and ready.

Altmetrics have many disadvantages when used for research evaluation. One of the typical challenges of using web indicators for research evaluation is that they lack quality control. Users are often anonymous, so there is no trail of evidence and they can easily be manipulated. Accidental manipulation might occur from publicity by the authors or publishers. Wouters and Costas (2012) argued that researchers could deliberately manipulate their usage indicators to improve their own impact scores. For example, lecturers can promote their own works to their students, knowing that an alternative indicator could be used to assess their research work. Zimmermann (2013) suggests that manipulation could be

reduced, if the evaluation is not of a high value-nature, by introducing honesty clauses or a degree of random or automatic checks on data for any sign of manipulation.

2.3.3 Validity assessment methods

As with all new indicators, evaluations altmetrics are necessary before they can be used with confidence. Articles can be mentioned to be criticised (Shema *et al*. 2012), or challenged for their non-scholarly contributions to research (Marcus & Oransky, 2011). Thelwall *et al*. (2013) argued that counts of tweets might not be indicators of public interest.

The standard way to partly authenticate the quality of any web indicator is to calculate the level of correlation with citation counts from the Web of Science (WoS) or Scopus (Sud & Thelwall, 2014). For example, early studies of different types of Web citations investigated whether they had a positive correlation with citation counts (Vaughan & Shaw, 2005; Kousha & Thelwall, 2007; Thelwall & Kousha, 2008). Correlations with citations are also a standard method in altmetrics research (e.g., Li *et al*., 2012) and for download counts (Schlögl, *et al*., 2014),

Correlation tests have also been used with other types of impact evidence, such as university rankings. The volume of commercial search engine searches for university names has been discovered to correlate with the ranking of the university (Tang & Thelwall, 2004; Vaughan & Romero-Frías, 2014). Hyperlink counts have also been correlated with university performance measures (Smith & Thelwall, 2002; Li *et al*., 2003; Stuart & Thelwall, 2005).

Normally a Spearman correlation test is better than a Pearson correlation test because citation data is usually skewed. When a correlation between an indicator and another source of research data is calculated, a positive result suggests that both sources reflect the same type of scholarly impact to some extent. The magnitude of the correlation is determined by the amount of the natural random data fluctuations as well as the extent to which the two sources reflect the same type of impact.

Since numerous studies have shown citation counts reflect scholarly impact in many fields, a positive correlation between citation counts and a web indicator suggests that the web indicator is also likely to correlate positively with research quality (Thelwall, 2016b). Nevertheless, it is technically possible for both citation counts and web indicators of the same articles to reflect positive correlations, but not be relevant to research quality. In addition, alternative indicators might have a zero correlation with citation counts if they reflect different types of research quality. Despite these possibilities, correlations with citation counts are a necessary first step to authenticate alternative indicators because of the lack of direct research quality evidence for journal articles. Statistically significant positive correlations also provide evidence that a web indicator is not purely random (Thelwall, 2016b).

The strength of a correlation partly reflects how close the relationship between the indicator and citation counts is (Thelwall (2016b). The correlation coefficient can be greatly reduced in strength if the sets of articles being compared are not in the same fields or years. Articles with

scores that are dominated by zeros or low numbers can reduce the values of the correlation and the strength of the correlation coefficient.

Other methods to evaluate indicators include content analysis of the web citation sources, interviews with those that created the citations or mentions, investigations of their use in decision making, and analyses of articles that have unusually high or low scores, compared to citation counts (Sud & Thelwall, 2014).

Although not a validity assessment, evaluating the proportion of articles with a non-zero score is an important practical step to gauge whether an altmetric provides enough data to be useful.

2.3.4 Validity assessment results

Many social media indicators have been assessed for their value as altmetrics, usually by comparing them with citation counts on the basis that the two should correlate positively even if an altmetric reflects a different type of research impact. Thelwall *et al.* (2013) investigated eleven altmetrics using data from Altmetric.com and found that they all had statistically significant associations with citation counts, but most were very low: Tweets (0.190), Facebook wall posts (0.050), Research Highlights (0.373), Blog mentions (0.201), Mainstream Media Mentions (0.088), Forum posts (0.033), Google posts (0.034), LinkedIn (0.009), Pinners (0.005), Question & Answer comments (0.048) and Reddit posts (0.062). The very low associations suggest that, in practice, most altmetrics have little value for individual journal articles, although they may still be useful for comparing sufficiently large groups of articles.

In a similar study, Costas *et al*. (2015) reviewed the presence of different altmetrics and their relationship with citations and journal indicators. A minority of the WoS articles investigated had been cited in Twitter (13%), Facebook (2.5%), Blogs (1.9%), Google+ (0.6%), and News outlets (0.5%). The proportion of articles with positive altmetric scores varied by subject category: Biomedical and health sciences (22.8%), Life and earth sciences (15.9%), Mathematics and computer science (5.4%), Natural science and engineering (9.0%), and Social sciences and humanities (22.5%). The authors concluded that altmetrics are most useful for recent publications and their presence is increasing over time and hence they may have increasing value to compliment citations for the analysis and evaluation of scholarly publications. This thesis covers only one of the social media metrics, the online social reference manager Mendeley readership counts.

Wouters and Costas (2012) showed that altmetrics are subject to many disadvantages when they are used as research indicators for formal or informal research evaluations. The authors argued that indicators acquired from the web are easy to manipulate and unsafe for most formal research evaluations.

Mas-Bleda and Thelwall (2016) used bibliometric and social media indicators to compare Spanish and UK research in eight subject fields with Scopus citation counts published in 2012. The study showed Mendeley has 80% coverage (articles having one or more Mendeley readers), compared to Twitter (34%), Wikipedia (26%), Syllabus mentions (1.2%), online

PowerPoint presentation citations (1.1%) and Patent citations (0.3%). In comparison, 80% of Spanish articles had Scopus citations, whilst 91% of the UK articles had at least one or more Scopus citations. The authors concluded that academic web development seems to be led by the English-speaking nations such as the USA and the UK, hence the assessment of data might be biased against non-English speaking nations compared to Scopus and the Web of Science.

2.5 Mendeley and other reference managers

Academic research that is read may be used in some way even if this use does not lead to a citation or a mention on the social web. Usage evidence for academic publications can be captured to some extent by counting readers in online reference management tools such as Mendeley, CiteULike and Zotero. Online reference manager tools generally allow users to save reference information from websites, and reference the online libraries as well as sharing this information with other users of the website.

Priem and Hemminger (2010) proposed that indicators derived from online reference managers such as Mendeley and Zotero and other social media tools (e.g. Twitter, Facebook and Wikipedia), could be useful to reflect the impact of articles earlier than citation counts. The authors suggested that for timely research evaluation, social media tools can be used instead of citation indicators. The above social reference sharing sites seem to have less users in most subject areas than Mendeley, as reported by Li *et al*. (2012). Most reference managers have no Application Programming Interface (API) that allow the collection of data for research evaluation. Nevertheless, Mendeley has an API that can report the number of users that bookmarked an article on the site, and this bibliographic tool can support scholars in managing their documents and citations.

During the second half of 2016, Thelwall (2017d) compared the average number of Mendeley readers to the average number of Scopus citation counts for 104520 articles from ten disciplines, and found between 0.1 and 0.8 of Mendeley readers per article in all fields when they appeared in Scopus for the first time. The author observed that most articles had no Mendeley readers when they were published. Hence, Mendeley readers could not be used to differentiate the average impacts of articles when first published. The author concluded that at the month of publication, Mendeley has more readers than Scopus citations per article, though the average is not enough for article-level evaluations for all subjects. The results of the study showed that there are disciplinary differences, with articles on Genetics being read the most and History being the least. The results also showed that Mendeley reader counts can be more useful than Scopus citations in terms of early evidence of scholarly impact for research evaluation, but neither Mendeley reader counts nor Scopus citation counts are large enough for differentiating, between the impacts of various articles, hence few of the articles had at least one Mendeley reader, when published.

Maflahi and Thelwall (2018) reported a weekly study of Mendeley readers of articles in six Library and Information science journals: Journal of Documentation; Journal of Information Science; Journal of Informetrics; Journal of the Association of Information Science; Library

and Information Science Research; and Scientometrics, to ascertain how fast Mendeley reader counts accumulate within the first year of publication. The study found that in all the six journals, readership counts experience a steady increase per article from the date of publication, hence Mendeley readers are confirmed to occur within weeks of publication. The study also demonstrated that articles that were registered by Scopus for Journal of Informetrics, Journal of the Association of Information Science, Library and Information Science Research, and Scientometrics, can attract citations for papers published with 'in press versions', except Journal of Documentation, or Journal of Information Science. Thus, the findings suggest that Mendeley readers do not appear immediately when an issue is published, but increase steadily when the online version of the article is made available.

2.5.1 Introduction to Mendeley

Mendeley (http://www.mendeley.com/) is a free academic social reference management website launched in 2008, often referred to as a social bookmarking tool. It is an academic social reference site where researchers can save and discover research papers. Mendeley has a desktop version for citing and managing PDF files, and an online version for sharing researchers' information with others and managing citations (Zaugg *et al*., 2011). Mendeley seems to be most popular among researchers in Life sciences, Chemistry, Mathematics and Computer Science (Gunn, 2013).

Several studies of Mendeley readership counts have shown that they are promising indicators of scholarly impact (Haustein & Siebenlist, 2011; Maflahi & Thelwall, 2016; Zahedi *et al*., 2013). They are amongst the most promising altmetric indicators (Li *et al*., 2012; Wouters & Costas, 2012), and appear much earlier than citations because they are less affected by publication delays (Bar-Ilan *et al*., 2012). For example, an article may be registered in Mendeley as soon as it is published. Mendeley also reveals the disciplines and nationalities of readers, giving more specific impact evidence (Thelwall & Sud, 2015). Several years ago, Mendeley had about 2.4 million users who had uploaded over 420 million documents (Gunn, 2013). Although Mendeley can be spammed, its large user base and positive results from previous analyses suggest that it does not currently suffer from a substantial amount of fake content.

The desktop version of Mendeley incorporates PDF management and annotation features, in addition to the common features for saving documents. The desktop version can operate independently or can be synchronized with the online server, enabling reference lists to be shared. Zaugg *et al*. (2011) observed that because of the social networking nature of Mendeley, researchers can use it to make their literature discovery tasks more efficient.

Mendeley has many features for users to store and manage bibliographical information. Whenever the desktop version is downloaded and installed, researchers can manually import bibliographical data files, such as BibTex, EndNote XML and RIS. The plugins in Mendeley allow PDF files to be imported, and automatically extract bibliographical information from the PDF metadata. The bibliographical information contains keywords and cited references for creating a bibliographic entry. The review folder holds the imported files until researcher verifies this bibliographic information. When there is missing metadata, this can be manually

corrected. Folders on the computer's hard drive, and imported PDF files placed into these folders, can be automatically monitored by Mendeley. Like other citation managers, bibliographical information can be saved by users from within most academic search engines such as ScienceDirect, Google Scholar, EBSCO and JSTOR, or can be saved directly to Mendeley (Zaugg *et al*., 2011). In addition, multiple citations can be organized by assigning tags to collections of folders, and information can be imported into the Mendeley library using other citation managers like Zotero, EndNotes, and RefWorks.

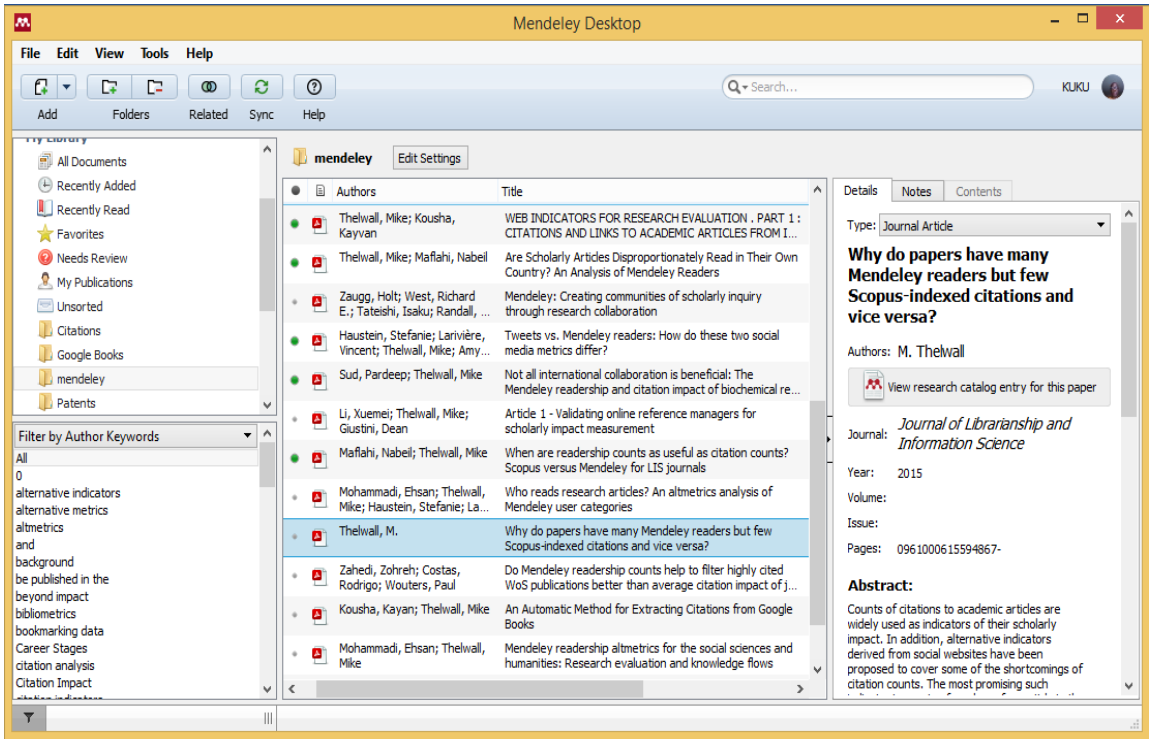


Figure 2.1 Mendeley Desktop, showing the library, filter, and the documents (Aduku K.J. Personal Mendeley Desktop, 2017)

Mendeley online has free document space of 500 MB. This allows researchers to perform functions such as synchronizing bibliographical information, notes, tagging information and retrieval of PDF files from their Mendeley desktop to Mendeley online library. Apart from the synchronization process Mendeley online has the capacity to allow researchers to edit their information and download PDF files.

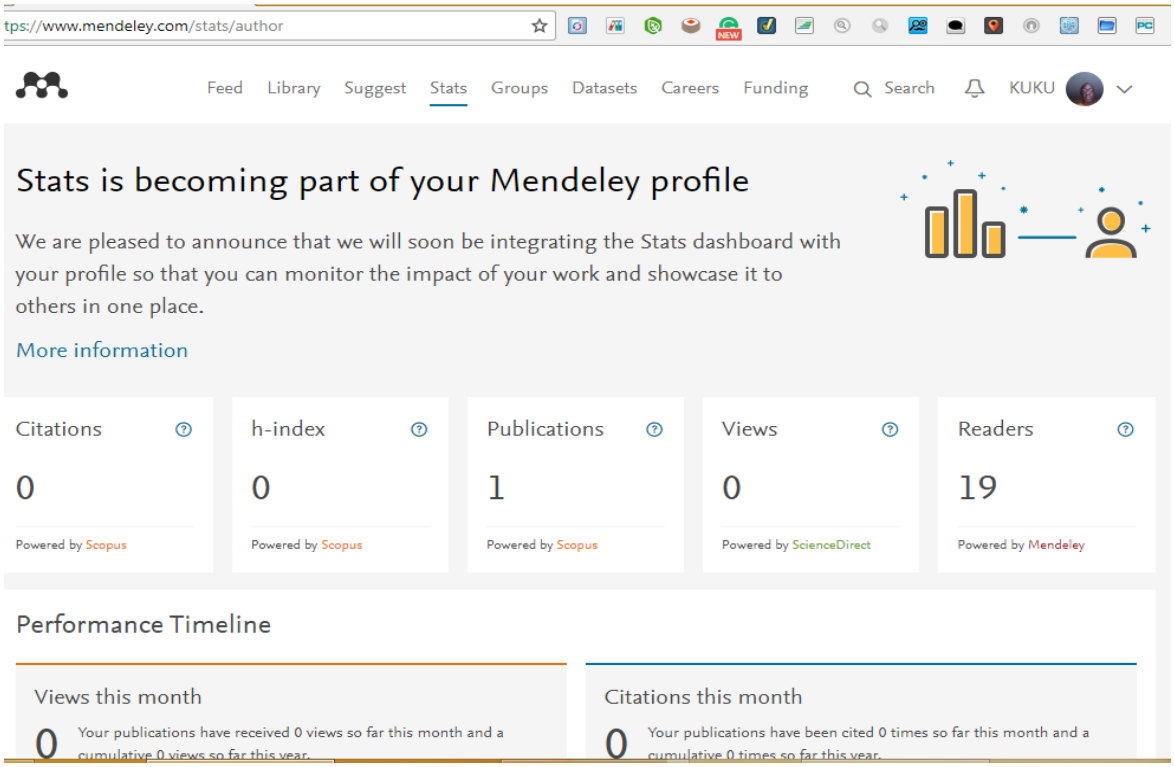


Figure 2.2 Online Mendeley user's statistics; citations, h-index, publications, views, and number of readers (Aduku K.J. personal online Mendeley, 2017).

Mendeley contains research statistics for individual documents (e.g., readership counts and counts by reader field, country, and job). It has recently also started to provide statistics for authors (see Figure 2.2).

Postdocs and PhD students are the main users of Mendeley also, Mendeley is been used by undergraduates, postgraduates, librarians and professors (Zahedi *et al*., 2013).

2.5.2 Correlations between Mendeley readers and citation counts

Several studies have found substantial positive correlations between Mendeley readership counts and citation counts. The first analysis of Li *et al*. (2012) investigated Mendeley and CiteULike for *Nature* and *Science* for 1,613 journal articles published in 2007. For Nature articles, the study found higher correlations between Mendeley user counts and Web of Science (WoS)/Google scholar citation counts (0.559 and 0.592 respectively), than between CiteULike and WoS/Google scholar citation counts (0.366 and 0.396 respectively). Also, for Science articles, there were higher correlations between Mendeley user counts and WoS/Google scholar citation counts (0.540 and 0.603 respectively), than between CiteULike and WoS/ Google scholar citation counts (0.304 and 0.381 respectively).

Mohammadi and Thelwall (2014) investigated the relationship between Mendeley readership counts and citation for articles indexed by WoS in the year 2008 across social sciences and humanities fields. They also found statistically significant positive correlations between Mendeley readership counts and citation counts. The values of the correlations varied across disciplines. Core sciences tended to have higher correlations compared to subject categories

within the Social Sciences and Humanities. For instance, citations between Mendeley readership counts and WoS citation counts in Psychology were (0.514), Linguistics (0.454), Religion (0.363), Literature (0.403), History (0.428), Philosophy (0.366), Information & Library Science (0.535), Business & Economics (0.573), Education and Education research (0.484) as well as Social Sciences and other subjects (0.403). The correlations for social science disciplines overall ($r = 0.516$, $p<0.01$) are higher than that for humanities disciplines ($r=0.428$, $p<0.01$). The correlations for Social Sciences disciplines were medium, varying from ($r=0.403$, $p<0.01$) to ($r=0.573$, $p<0.01$) for Business and Economics. Amongst humanities disciplines, Religion and Philosophy have the lowest correlations ($r=0.363$, $p<0.01$ and $r=0.366$, $p<0.01$) and Linguistics has the highest correlation ($r=0.454$, $p<0.01$).

Haustein and Larivière (2014) found moderate correlations between readers and citations, suggesting that Mendeley readers captured usage patterns to different extents for Web of Science areas of research. For example, Spearman correlations indicate similar readership and citation patterns in General Biomedical Research ($\rho=0.689^{**}$), and Embryology ($\rho=0.649^{**}$) and but much less in Veterinary Medicine ($\rho=0.236^{**}$), Social Studies of Medicine ($\rho=0.281^{**}$) and Psychoanalysis ($\rho=0.137$). Similarly, a study of scientific publications in Mendeley by Zahedi *et al*. (2015) across 5 major fields of science also reported correlations between Mendeley and citation counts. Social sciences and humanities had the highest correlation ($p=0.614$), Natural Sciences and Engineering ($p=0.597$), Life and Earth Sciences ($p=0.578$), Biomedical and Health Sciences ($p=0.553$) and Mathematics and Computer Sciences also had moderate correlations ($p=0.457$). From the different types of users, PhD students had the highest correlations, whereas, Students, Post Doctorates, Researchers, Professors and other Professionals had decreasing correlations, in that order. Librarians and Lecturers had the lowest correlations.

Mohammadi *et al*. (2015) reported Spearman correlations with 2008 WoS articles in Clinical Medicine (0.463), Engineering and Technology (0.327), Social Sciences (0.456), Physics (0.308) and Chemistry (0.369). These correlations are all positive and moderate. The study also calculated correlations between Mendeley readership counts and citation counts for different types of user. Although small correlations were seen between undergraduates and other professionals, there were higher correlations for Clinical Medicine and other professions when compared to other disciplines.

Thelwall and Wilson (2016) investigated a sample of 332,975 articles from 2009 in 45 medical fields in Scopus, to assess correlations between citation counts and Mendeley readership counts. The citation counts correlated strongly (0.7) with Mendeley readership counts and about 78% of articles had at least one reader. The correlation was weak for a small sample of the Drug Guides category (0.37) but was still positive. The study discovered similar correlations when articles that were not found in Mendeley were assumed to have zero readers, compared to these articles being treated as missing data and removed before calculating correlations. The findings were restricted to a single year. The results provide confidence that Mendeley readership counts could be used as evidence for citation impact in all areas of medical research, except perhaps for drug guides because of its non-significant result.

In the most detailed study so far, Thelwall (2017d) compared Mendeley reader counts with citation counts for journal articles in 325 Scopus narrow field categories from 2012, to ascertain whether Mendeley reader counts could be useful impact indicators in all fields. The study revealed strong positive correlations (0.671) overall between Mendeley readership and Scopus citation counts. Although in some fields the correlations were as weak as 0.255, this might be due to technical problems such as: low proportions of readers per paper, errors due to Scopus indexing, and that articles in some disciplines having no DOI, making it difficult to locate them in Mendeley. Low correlations can also be caused by topics of interest within countries that rarely use Mendeley. The study suggests that if Mendeley readership counts are carefully used for early research impact, evidence can be accepted across all the fields as impact indicators.

Table 2.1 Evidence of correlations between Mendeley readers and citations for journal articles.

Study	***Main subject areas covered by the study***	***Dataset (No. of articles or years)***	***Evidence for correlation association between Mendeley and citation indicator***
Thelwall and Wilson (2016)	45 Medical Fields	332,975 articles from 2009	Moderate (ranging from 0.4 to 0.7)
Li *et al*. (2012)	Nature and Science	1,613 articles from 2007	Moderate (ranging from 0.5 to 0.6)
Mohammadi and Thelwall (2014)	Social Sciences and Humanities	2008	Moderate (ranging from 0.4 to 0.6)
Haustein and Larivière (2014)	Biomedical Research, Clinical Medicine, Health and Psychology	1.2 million documents	Weak/Moderate/Strong (ranging from 0.1 to 0.7)
Zahedi *et al*. (2015)	Biomedical & Health Sciences, Natural Sciences & Engineering, Life & Earth Sciences, Social Sciences & Humanities and Mathematics & Computer Science	1, 196,421 articles from 2011	Moderate (ranging from 0.4 to 0.6)
Mohammadi *et al*. (2015)	Clinical Medicine, Engineering and Technology, Social Sciences, Physics, and Chemistry.	480,979 articles from 2008	Moderate (ranging from 0.3 to 0.5)
Thelwall (2017d)	325 Scopus sub-fields	1,257,148 articles from 2012	Moderate/ Strong (ranging from 0.25 to 0.67)

2.5.3 Field differences in the use of Mendeley

For 2008 WoS journal articles, in Clinical Medicine (71.6%), Engineering and Technology (34.8%), Social Science (46.8%), Physics (31.4%) and Chemistry (33.7%) had at least one Mendeley reader (Mohammadi *et al*., 2015). The authors observed that the correlation

strengths vary by occupation across the research disciplines. The study revealed that about 66% of correlations decreased within the readership occupations available, in comparison with the records of 100% available from the readership occupations. In addition, the study showed that all correlations were lower for papers with 100% readership occupations available. The authors suggested an outcome of the research was the possibility that these papers are the least cited papers with the lowest total number of readers; hence the correlation test is less powerful for them because the numbers were smaller.

Mohammadi *et al*. (2013) surveyed 860 Mendeley users and found that most researchers bookmarked papers in their Mendeley libraries with the intention either to read them or for future reference. From the survey respondents, 27% were PhD students, 26% were postdoctoral researchers, 14% were student assistants, 13% were associate professors, and 11% were full professors, while 6% were other professionals. Masters and undergraduate students were represented in low proportions, 3%, and 1% respectively. 78% of respondents had a personal library in Mendeley and 87% used Mendeley as a reference manager. 30% of researchers used Mendeley for searching academic publications, 25% used Mendeley to share publications and 15% used Mendeley for social networking. About 85% bookmarked articles in Mendeley to cite them in their publications, while 50% recorded them for professional reasons, 25% for teaching purposes and 13% for educational activities.

In contrast, a survey of 146 open group participants on Mendeley users by Jeng *et al*. (2015) showed that the majority of Mendeley users have a background of higher education; doctoral students, post-doctoral fellows and graduate students. The study observed that most users of Mendeley were from the fields of computer & information science and biomedical sciences, the fields of social sciences, education and psychology are more recent users of Mendeley.

Similarly, Thelwall (in press) analysed twelve narrow Arts & Humanities subcategories from Scopus between 2007 to 2017, to indicate the value of research using Mendeley reader counts. The study showed medium to high correlations between Mendeley readership counts and Scopus citation counts for: History (0.622); Language and Linguistics (0.815); Archeology (arts and humanities) (0.707); Classics (0.384); Conservation (0.729); History and Philosophy of Science (0.779); Literature and Literary Theory (0.382); Museology (0.727); Music (0.634); Philosophy (0.634); Religious Studies (0.512); Visual Arts and Performing Arts (0.477). However, these findings contrast with the previous findings by Mohammadi & Thelwall, (2014) where five Web of Science humanities fields (Linguistics); Philosophy; History; Literature; and Religion) were investigated and found low to medium (0.2-0.3) correlations. However, the author concluded that Mendeley may be more extensively used in the field of humanities in 2017 than it was in 2014, considering the higher percentage of readership counts and higher correlations.

Table 2.2 Proportions of articles with Mendeley readers from prior studies.

Study	**Main subject areas covered by the study**	**Dataset (No. of articles)**	**Proportion of articles with Mendeley readers**
Zahedi *et al*. (2014a)	Impact Story (IS) altmetrics (Mendeley, Twitter, Wikipedia, and Delicious	19722 articles	62.9%
Haustein and Larivière (2014)	Biomedical Research, Clinical Medicine, Health and psychology	1.2 million documents	65.9%
Zahedi *et al*. (2014b)	Biomedical & health sciences, Natural sciences & engineering, Life & earth sciences, Social sciences & humanities and mathematics & computer science	1,107,917 articles from 2011	Academic status (Ranging from 1.3% to 42%) and Subject areas (Ranging from 6% to 38%)
Zahedi *et al*. (2015)	Biomedical & health sciences, Natural sciences & engineering, Life & earth sciences, Social sciences & humanities and mathematics & computer science	1,196,421 articles from 2011	Subject areas (ranging from 8% to 37%)
Mas-Bleda *et al*. (2014)	Social web sites	45 European countries. From 2012	Low (Ranging from 4% to 6%)
Mohammadi *et al*. (2015a)	Clinical Medicine, Engineering and Technology, Social Science, Physics and Chemistry	480,979 articles from 2008	High (Ranging from34% to 72%)
Mohammadi *et al*. (2015b)	Mendeley Readers	2013 survey of 860 Mendeley users	High (55%)

2.5.4 Mendeley users

Zahedi *et al*. (2014a) used the Impact Story (IS) website to investigate the presence and distribution of altmetrics (Mendeley, Twitter, Wikipedia and Delicious) across fields, showing that 62.6% of 19722 journal articles had at least one Mendeley reader. Wikipedia (1.4%), and Delicious (0.3%) are all rare for journal articles, in comparison. Similarly, Haustein and Larivière (2014) investigated Mendeley usage statistics for a sample of Biomedical Research, Clinical Medicine, Health and Psychology, comprised of 1.2 million journal articles. They found that 65.9% of the articles had at least one Mendeley reader. The study found that 33.2% of users are PhD students, 17.7% Postgraduate students, 11.1% Post Doctorates, 7.2% academic institution researchers, 7.0% other professionals, 5.5% assistant professors, 4.7% bachelor students, 4.5% non-academic institution researchers, 4.0% professors, 3.6% associate professors and 1.4% librarians.

Haustein and Larivière (2014) investigated correlations between citation counts and reader counts for different reader occupations and fields. The highest correlations with citations can be found in Clinical medicine, academic and non-academic researchers, whilst the highest Spearman values in health are obtained from PhD students, academic and non-academic researchers and Postdocs. In Clinical Medicine and Health, Bachelor students' citation counts had low Spearman values (0.236), whereas in Biomedical Research and Psychology, their correlations (0.683) were slightly higher. In addition, the authors argued that low citation and reader counts are due to the use of scientific papers for practical applications. Similar results were found in a study of 2008 WoS journal articles (Mohammadi *et al*., 2015). The highest correlations were from full Professors, Assistant Professors, Post Doctorates and PhD students, while the lowest correlations were from undergraduates, other professionals and librarians in all disciplines, from all the three datasets.

Zahedi *et al*. (2014b) showed that PhD students form almost half of all Mendeley readers (42%), while Lecturers and Librarians are the least common user types with 16.2% and 1.3% respectively.

2.5.5 Mendeley reader count outliers

For Mendeley reader counts citations to be used as scholarly impact indicators with confidence, citation-reader anomalies need to be explored. Thelwall (2017b) provides some reasons why articles can be anomalies in the sense that they have relatively many (or few) Mendeley readers for their number of citations. These reasons include the following, all of which may also apply to conference papers.

- Many students do not produce academic work that should be indexed in Scopus but read articles.
- Professionals may read, but may not necessarily publish, articles
- Multidisciplinary articles may attract more readers from different fields than citations

- Topics may attract casual readers (i.e. readers of articles based on their captions, not necessarily for any academic important)
- Topics may be of interest to readers that mainly come from countries that rarely publish in Scopus journals.
- Academic communities may not use Mendeley due to limited internet access.
- Some professions, such as hospital doctors, rarely use Mendeley due to working practices but, may publish articles to be indexed in Scopus.
- An update of an article may be registered by Mendeley users while authors may cite the original.

2.5.6 Other important studies about Mendeley

Multiple aspects of scholarly communication can be investigated using readership data. Haustein and Siebenlist (2011) showed that Mendeley readership data can be used for evaluating academic journals as an alternative to citations. Heck *et al*. (2011) proposed the use of data from Mendeley readership and data from CiteUlike for information retrieval. Borrego and Fry (2012) observed that data from BibSonomy can provide insights about uses of scholarly information. Zahedi *et al*. (2013) investigated Mendeley readers based on their career stages, disciplines and countries. They found that PhD students read more recent papers, and papers that have higher impact. Thelwall and Maflahi (2015) showed that Mendeley reader data can be used to track the relationship between the origin of the authors of the articles and the countries of their readers. Also, Kraker *et al*. (2014) showed that Mendeley data can be used for mapping a discipline.

2.6 Google Books citations

There is substantial evidence that books are the primary research outputs in some social sciences, arts and humanities fields (Huang & Chang, 2008). This is reflected in their dominance in reference lists for publications in these fields. For example, for British history journal articles published between 1968 and 1969, 34.1% of the 7,127 citations were to books, followed by journal articles with only 21.5% (Jones *et al*., 1972). For articles published between 1965 and 1974, in three core music-research journals, 38.8% of their references were to books, while 18.3% were to journal articles (Baker, 1978). For music theses submitted between 1975 and 1980, 58% of their citations were to books, whereas only 29.8% were to journal articles (Griscom, 1983). Budd (1986), reviewing 253 American literature journal articles, found that 64% of their citations were to books and only 23% were to journal articles. A study of philosophy and political science doctoral dissertations by Buchanan and Herubel (1993) found that 81.3% of their citations were to books, followed by 13.3% to journal articles. The investigation of Chung (1995) of 5,302 citations from classification-related publications between 1981 and 1990, showed that books and book chapters formed 51.3% of the citations, whereas 37.8% citations were to journal articles. The work of Hider (1996) on the British anthropology journal, *Man: Journal of the Royal Anthropological Institute*, showed that over the years the most commonly cited publications

were books, followed by journal articles. Similarly, citation studies by Knievel and Kellsey (2005) on humanities literature found that citations to books form an average of 74.3% of the total citations across the fields, while journal articles formed 25.3% and electronic resources accounted for 0.3%.

From a different perspective, Nederhof and Noyons (1992) investigated the citations received by the publications of general linguistics and general literature research groups, finding that their books and book chapters received the most citations in total, with few citations to their journal articles. Thus, it is clear that books are important scholarly outputs in some fields and it is therefore important to extract citations *from* books.

Google Books (GB) is the largest free full text index of books. Books play a significant role in scholarly communication, especially in social sciences and humanities (Kousha & Thelwall, 2009). Although Google Books does not provide a citation index, the use of Webometric Analyst makes it possible to query book citations by means of entering the book title as a phrase search, with author name and year of publication added to narrow down the search results. The results may need to be manually filtered to identify correct matches, however, at least with the early version of Webometric Analyst (Kousha & Thelwall, 2015a).

Google Books citations have positive Spearman correlations with WoS citations in sciences (0.15 to 0.35; except for computing with 0.71), social sciences correlations are 0.41 to 0.59, and 0.36 to 0.65 in the humanities. For citations to academic monographs rather than journal articles, Google Books citations in humanities are 1.4 to 3.2 times as frequent as Scopus citations (Kousha *et al*., 2011). For a non-English language developing nation, Abrizah and Thelwall (2014) assessed books published by Malaysian university presses between 1961 and 2012, showing that more had Google Books citations than Google Scholar citations. Similarly, Kousha and Thelwall (2015a) on the proportion of Google Books citations in most Book Citation Index (BkCI), found that by 2008, most monographs had at least one Google Books citation in the following fields: 92% arts and humanities, 85% social sciences and 70% science and medicine. Kousha *et al*. (2016) in 2015 found that the following fields had at least one Google Books citation: arts and humanities (81-88%), social sciences (79-87%), science (53-70%), Medicine (42-80%) and Engineering (53-69%). Google Books citations to books were almost as common as WoS citations to books in Social Sciences.

Kousha and Thelwall (2015b) found low correlations between Google Books citations and numbers of Amazon reviews (0.171) for the best-selling books. For 15 disciplines, there are low to medium significant correlations between Google Books citation counts and Mendeley reader counts. The authors concluded that online reviews might reflect a wider audience both inside and outside academia, in terms of capturing broader educational or cultural activities that cannot be manipulated in research evaluation. No prior studies have investigated Google Books citations to conference papers, however.

2.7 Google Patents citations

Google Patents (www.google.com/patents) is a search engine launched in 2006 that indexes patents from multiple patent offices. Orduna-Malea *et al*. (2017) showed that Google Patents'

website indexes full text patents and patent applications from several sources such as: United States Patent and Trademark Office (USPTO), European Patent Office (EPO), World Intellectual Property Organization (WIPO), Deutsches Patent und Markenamt (DPMA), Canadian Intellectual Property Office (CIPO) and China's State Intellectual Property Office (SIPO). The study also revealed that URL citations can also be extracted from the Google Patents website. A citation from a patent to an academic paper can be used as an indicator of commercial value of the cited paper because the task of a patent it to protect the commercial value of an invention. Google Patents is a useful source of patent citations because it is free. Although it does not have a citation index, it is possible to search for academic papers with Google Patents and, if found, check that the papers are cited in a patent. A few studies have done this for journal articles.

Kousha and Thelwall (2015c) analysed journal articles published between 1996 and 2012 in Scopus subject categories for sixteen applied science and engineering fields. The coverage (number of journal articles with at least one citation) of Google Patent citations were all below 11%: Biochemistry & Molecular Biology (5.4%), Biomedical Engineering (10.1%), Biotechnology (9.2%), Chemical Engineering (2.9%), Computer Science (5.9%), Control & Systems Engineering (3.9%), Electrical & Electronic Engineering (5.6%), Energy Engineering (2.2%), Environmental Engineering (2.7%), Food Science (5.5%), Industrial & Manufacturing Engineering (2.5%), Mechanical Engineering (1.9%), Pharmacology & Pharmaceuticals (6.8%), Physics Instruments & Instrumentation (2.8%), Polymer Science (4.6%) and Surgery (2.9%). There were weak (0.05 to 0.36) positive correlations amongst Scopus citation counts across all the applied science and engineering fields analysed. Thus, the study suggests that a small minority of articles are cited by patents in most academic fields, even if they are of an applied nature.

The Google Patents search engine does not allow automated queries but can reveal patent citations through automated Bing queries of the Google Patents website. It is therefore possible to extract patent citation counts for use as a commercial impact indicator, although a small charge is currently required to submit automated queries via Bing, beyond the initial free amount (currently 10,000) (Kousha & Thelwall, 2015c). No previous study has analysed Google patents citations to conference papers, however.

2.8 Wikipedia citations

Wikipedia is a free web-based edited encyclopaedia that was launched on 15 January 2001, and is written in collaboration by volunteers in many languages. Wikipedia provides a widely-used overview of many academic fields, often referencing journal articles and books to justify its content. Previous studies have shown that these citations can in turn be used to help assess the knowledge transfer impact of the cited articles and books. Wikipedia has experienced exponential growth since 2002 (Kittur *et al*., 2007). In the USA, over a third of college students had used Wikipedia by 2013, despite concerns about its quality and reliability (McKerlich *et al*., 2013; Aibar *et al*., 2015; Knight & Pryke, 2012; Soules, 2015). Bould *et al*. (2014) on Wikipedia as a source of credible information showed that 70% of junior physicians consult Wikipedia to acquire health information. Due to its popularity, its

coverage, currency, accuracy and readability have all been investigated (Mesgari *et al*., 2015). Previous studies have analysed Wikipedia contributors (Jullien, 2012; Yasseri & Kertész, 2013) and used it as raw data for text-mining (Medelyan *et al*., 2009).

The unique openness of Wikipedia is a key strength in attracting new contributing users (Kittur *et al*., 2007). The reliability of information in Wikipedia has always been a controversial subject (Waters, 2007; Jaschik, 2007; Chen, 2007; Carleton College, 2007; Ver Steeg, 2008), with some comparing its contents to encyclopaedia that centrally monitor and edit their content, and others creating doubt on the trustworthiness of its contents. Chen and Roth (2012) claimed that researchers are making effort to understand the way and manner Wikipedia articles are edited, and the conditions under which collective editing will lead to more productive and higher quality articles.

Between 2001 and 2010, English language Wikipedia accrued about 3.3 million articles, and in January 2010, Wikipedia attracted about 362 million viewers globally (Heilman *et al*., 2011). Wikipedia was the sixth most popular internet website and hosted more than seven million media files in the Wikipedia central repository. Jimmy Wales, the founder of Wikipedia, argued that a small number of users did most of the work on Wikipedia. He stated that as of December 2004, half of all Wikipedia edits were done by 2.5% of the registered Wikipedia users (Wales, 2005 in Kittur *et al*., 2007). Kittur *et al*. (2007) confirmed that since the creation of Wikipedia, a small group contributed most of the work, but claimed that the distribution of the work had shifted to the common users with a corresponding decline in the influence of the elites. This paper did not specify whether the work was measured by edits or changes in the content. A content analysis showed that the elites added more words per edit, while novice users removed more words than they added.

Suh *et al*. (2008) argues that the concept of Wikipedia can be viewed as "the academic process for ascertaining the truth". They stated that information contained in Wikipedia is transparent, giving everyone access to edit the content, making necessary corrections to an incorrect source of data and examining any opposing point of view side by side. There are many empirical studies on the credibility and quality of Wikipedia's articles as a source of scholarly information (Wilkinson & Huberman, 2007; Halavais & Lackaff, 2008; Royal and Kapila, 2008). Park (2011) found that there is an exponential increase in scientific studies concerning Wikipedia. Information from Wikipedia should not be used for primary research, however, except when Wikipedia is used as the topic of the research (Wikipedia: Researching with Wikipedia). Despite this, Tomaszewski and MacDonald (2016) showed that Wikipedia citations in peer reviewed publications have increased between 2002 and 2015.

Despite limitations and challenges for using Wikipedia to ascertain academic source of information, Wikipedia citations can reflect scientific or research impact, educational impact in some popular science entries, and cultural impact through biographies or history entries. Wikipedia citations can reflect general informational impact and book assessment impact, especially in the Arts and Humanities (Kousha and Thelwall, 2017).

Thomas *et al*. (2013) have demonstrated how reliable the source of Wikipedia to scientific journals, where articles from the Encyclopaedia Britannica were compared to 42 scientific articles similar accuracy levels were found. Wikipedia's articles sometimes cite academic research as evidence for reported facts. A citation from Wikipedia may be taken as evidence that the cited paper has created knowledge that may be useful to the audience of Wikipedia. Citations from Wikipedia seem to reflect educational or general informational impact, although this depends on the subject area or topic of the cited document (Kousha & Thelwall, 2016).

Several studies have analysed citations from Wikipedia articles to help assess their value and any biases. Nielsen (2008) revealed bias in Wikipedia citations towards some prestigious journals. For example, the Astrophysical Journal had 424 citations, Astronomical and Astrophysics had 154 citations, Icarus International Journal of Solar Studies 147 citations and Astronomical Journal 93 citations. Some of the medical journals such as The Lancet had 268 citations, JAMA had 217 citations, British Medical Journal 187 citations, and Annals of International Medicine 104 citations. The study revealed that astronomy journals received more Wikipedia citations compared to citations in medical journals, suggesting a field bias. In addition, some journal articles received many citations due to the concerted efforts of their respective authors. For example, The Australian Botany Journal has 101 citations due to the efforts of one author, Banksia. Journals from computer and Internet-related fields have few citations. For instance, Communications of the ACM has only 34 citations despite being a flagship journal.

In the health field, Laurent and Vickers (2009) showed that the English Wikipedia is an important source for health-related information in comparison to MedlinePlus, NHS Direct, and the National Organisation of Rare Diseases. Medical keywords were searched as queries through special software to measure Wikipedia ranking, and number of pages reported in internet search engines (Google, Google UK, Yahoo, and MNS). The study concluded that English Wikipedia was more often in the first ten Google search results than any medical website. Practicing physicians use resources from Wikipedia to provide medical care, and junior physicians used Wikipedia almost every week, more often than all other medical websites (Heilman *et al*., 2011). Wikipedia receives about 150 million-page views on medical articles every month, compared to other sources of medical articles that receive only 60,000 views per month.

Citations to Wikipedia from academic research have also been investigated, although many journals do not allow this. Wikipedia is cited by some English Medical Science journal articles. From PubMed and Medline, 1,433 articles from 1,008 journals citing 2,048 Wikipedia articles. 31.6% cited for main definitions and 23.5% cited processes definitions (Bould *et al*., 2014). Between 2005 and 2009, articles in chemistry journals also cited Wikipedia at least 370 times (Brazzeal, 2011). Overall, there has been an exponential increase of Scopus publications citing Wikipedia: in August 2015, about 61,135 Scopus publications cited at least one Wikipedia article. In comparison, only 7,849 Scopus publications cited Encyclopaedia Britannica in the same year (Kousha & Thelwall, 2016).

Wikipedia citations to academic articles and books can be extracted automatically, as demonstrated by Kousha and Thelwall (2017). Wikipedia citations can reflect scientific or research impact, educational impact in some popular science entries, and cultural impact through biographies or history entries. Wikipedia citations can reflect general informational impact and book assessment impact, especially in Arts and Humanities.

In terms of books, Halfaker and Taraborelli (2015) showed that ISBN, PubMed, DOI and ArXiv identifiers in English Wikipedia can be used to identify cited books. From all their matching records 35% cite books and 2% cite academic journals, suggesting that books are an important source of information for Wikipedia. Kousha *et al*. (2016) investigated 15,928 BkCI academic books published between 2008 and 2010 and showed that by 2015, about half had at least one Wikipedia citation. The proportions with at least one Wikipedia citations were: Arts (58-61%), Humanities (48-54%), Social Sciences (30-39%), Medical Sciences (18-34%), Science (23-35%) and Engineering (18-37%). The authors claimed that Wikipedia citations are useful indicators for academic books on the basis that they are relatively common, although the citations could be influenced by authors adding them to Wikipedia.

Although it is known that Wikipedia cites books and journal articles, no previous studies have analysed citations from Wikipedia to conference papers.

2.9 Impact assessment for conference papers

One venue for presenting research findings is at a research conference. Often, computer scientists and engineers prefer to publish in conferences than journals because they are more timely (Patterson, 2004). In support of this, Goodrum *et al*. (2001) showed that conference papers in some research fields accrued early citations, more cutting-edge recognition and were more cited than journal articles.

Conference proceedings represent an important part of the published literature in engineering (Glänzel *et al*., 2006). Goodrum *et al*. (2001), Visser and Moed (2005) and Butler (2008), all also argue that conference papers can be more important than journal articles in the transmission of knowledge for computer sciences and other related engineering fields.

Zhang and Jia (2013) investigated the policies of republishing many computer science journals which had already been published in conference proceedings. The study revealed that most journal editors would not like to republish such papers except in special circumstances. Some editors would accept publishing the paper after considerable additional new ideas from the original version. The authors observed that many editors would specify the amount of content that should be new, with 30% being mentioned often. Some sections of the text might still look like the original paper, with the proper citing of the original publication.

Vrettas and Sanderson (2015) examined a list of 20,712 journal articles and 1,952 conference papers in 2010 for Australian academics and found that 76% of the conference papers were from Computer Science while engineering was found to be the next largest research field with 12%. The authors examined the percentage of conference grade A (75 to 100%) to C (50

to 59%) and found that Computer Science fall into the lower grade of conferences B (60 to 74%) or C (50 to 59%), while Engineering disciplines were found in the highest-grade A (75 to 100%). The authors argued that the compilation of the engineering list might be based on the best conferences, whilst that of Computer Science appeared to have encompassed a wider selection.

Vardi (2009) argued that from the early 1980s, the fields of computer science and engineering were dominated by conference publications. The author claimed that between the 1980s and 1990s there was uncertainty in the research community because of pressure from promotion and tenure of committees comparing conferences with journals. According to Halpern and Parkes (2011) claim that the quality of computer science research was being assessed through publication in selective leading conferences.

In computing, many conferences employ rigorous peer review and have low acceptance rates (Terry, 2014). Strict limits are placed on the number of accepted papers and papers are rejected from which the community could benefit. Rejected papers could be resubmitted many times to different conferences before these papers are eventually accepted, or the authors give up in frustration. Good papers may have their publication delayed to the detriment of the research community, while poor papers may receive little attention and might not get the constructive feedback necessary to improve the quality of the submitted paper. Terry observed that the benefits of accepting any reasonable conference submission for publication might lead to knowledge contribution and reduce low acceptance rates. The author suggested that research results get published in a timelier manner, therefore reviewers should focus on providing constructive feedback.

An early study by Drott (1995) of conference papers from the 1987 Proceedings of the Annual Meeting of the American Society for Information Science found that of the 32 papers, only 13% were subsequently published as journal articles. Drott (1995) further observed that conference proceedings have three specific functions. First, young researchers can improve their papers through feedback from older researchers before final submission for publication. Second, they can engage experienced researchers in discussions in the field of their research, to have a fresh focus on their research area and to seek peer expertise. Third, they may discard information about their research that seems too ambiguous to be included in an article, reports or theories. Becher and Trowler (2001) argue that conference papers are starting points for journal articles or books in many disciplines.

Some conference papers are indexed by Scopus or WoS. A study conducted by Kademani *et al*. (1999) during 1999-2007 on the coverage of conference papers of BARC scientists and engineers showed that only 8.5% had been covered in *Scopus*. The average number of citations per paper was 1.81. In addition, the author observed that multi-authored papers received more citations than single authored papers. Similarly, the study of Arrive *et al*. (2004) on conference publications presented at the 1995 Radiological Society of North America, found that 33% of selected abstracts led to articles published in Medline-indexed journals.

A study of Natural Science and Engineering compared to Social Sciences and Humanities for conference papers indexed in Thomson Scientific databases by Lisée *et al*. (2008), showed that in Natural Sciences and Engineering (NSE) 19.6% of Computing conference papers had citations, compared to Electrical Engineering & Electronics (13.1%), Civil Engineering Had (11.5%), Nuclear Technology (11.2%), Miscellaneous Engineering & Technology (10.3%), Aerospace Technology (8.9%), Industrial Engineering (8.5%), and Mechanical Engineering (8%). Non-engineering fields included Acoustics (7.7%), Psychology Human Factors (7.6%), Library & Information Science (7.3%) and Optics (6%). There were lower rates in the Social Sciences and Humanities (SSH): Ergonomics (7.6%), Transport Studies (4.8%), Information Science and Library Science (3.3%), Demography (2.6%), Education Research (2.6%), Language and Linguistics (2.5%), Education (2.4%) and Environmental Studies (2.1%).

2.10 Summary

The above review has highlighted the need to replace conventional citation impact for research evaluation because publications that are not cited may still have an impact and other types of research activities may also have had some impact on the scientific community or on public. For example, articles can be cited by publications which are not indexed by conventional citation databases (Books, Patents and Encyclopaedia entries), and can be used for wider scholarly reasons which cannot be traced through citations such as for educational applications as well as for applications outside the academia. Moreover, citations can only be counted a long time after the research has been conducted.

The prior literature has shown that it is possible to generate a wide range of alternative indicators from the social web and general web, many of which have been investigated for journal articles. It also demonstrates that correlation tests are the standard method to assess new indicators but that other methods are also necessary to give a fuller picture of their value.

The literature shows that Mendeley reader counts are promising indicators of early scholarly impact since they appear much earlier than citations because they are as less affected by publication delays. The literature also suggests that Google patents citations (technological impact), Google Books citations (book-based impact or humanities impact) and Wikipedia citations (informational impact) can all reveal non-academic impacts for some types of scholarly outputs in some fields.

Chapter 3: Research Questions

This thesis aims to fill knowledge gaps in the literature, all of which concern the use of altmetrics for conference papers in engineering fields that value conferences. As discussed above, conferences are important in some fields and can be more important than journals. Despite this, no prior altmetric research has focused on conference papers but has instead investigated journal articles or books. Thus, if altmetrics are to be used in academia then it is important to assess altmetrics for conference papers so that they can be used on all the core research outputs as far as possible.

3.1 Choice of fields to investigate

There is no record of which fields conferences are most important because this is tacit information, known to field practitioners but rarely formalised through mass surveys or official pronouncements. Nevertheless, the above review shows that they are known to be important in engineering-related aspects of computer science (but probably much less important in theoretical aspects of computer science). For example, the World Wide Web conferences have rejection rates above 95% as does the key Information Retrieval conference, SIGIR (Association of Computer Machines Special Interest Group on Information Retrieval). Thus, this thesis investigates two applied computer science fields within Scopus: *Computer Science Applications* and *Computer Software*. Scopus was chosen in preference to the Web of Science for its wider coverage of conference papers and narrower subject categories.

Conferences also seem to be important in other engineering-related fields, although it is not clear whether they are ever regarded as more important than journals in any general engineering fields. Thus, this thesis also investigates general engineering fields: *Building & Construction Engineering* and *Industrial & Manufacturing Engineering*.

At least one important conference-based field is not analysed in this thesis. Computational linguistics is dominated by conferences and the most prestigious venue is the Association of Computational Linguistics (ACL) annual conference. This field was not chosen because it is relatively small and does not have its own Scopus category.

Thus, the scope of this article in terms of fields is four engineering fields in which conferences are important: Computer Science Applications; Computer Software; Building & Construction Engineering; and Industrial & Manufacturing Engineering.

3.2 Choice of altmetrics to investigate

Four altmetrics were chosen to be investigated in this thesis. Mendeley readers was chosen because they are the most robust indicator found so far for journal articles, and so are an obvious choice. As the above review shows, they are less used for books, and so it is not clear whether they would have as much value as journal articles.

Patent citations were chosen to investigate because patents are an indicator of applied value and engineering research is closely related to building things (e.g., computers, software,

buildings, machines), often with potential commercial value. They outputs may be directly sold or contribute parts of products that are sold.

Wikipedia citations are potentially relevant to engineering fields because encyclopedia incorporate records for many engineering outputs and so could potentially draw upon engineering research to describe "how things work" or widely adopted products.

Finally, Google Books was chosen because conference proceedings are often published as books with ISBNs that may be indexed in Google books. Evidence for the importance of Google Books searches for 1000 books in the UK Research Assessment Exercise in 2008 show that citations from Google Books are 1.4 times more common than citations from Scopus articles. Google Books has much greater coverage of books than WoS and Scopus but does not have a citation index. Because of its huge size and the possibility to extract citation counts from it, Google Books is a promising new source of citation evidence (Kousha, Thelwall & Rezaie, 2011). It may be particularly useful for conference papers in engineering fields because journal articles are less central to these areas and so traditional journal-based citation indexes, such as Scopus and Web of Science, may not serve them well. In general, there is substantial evidence that Google Books citations reflect scholarly impact. Thus, citations from Google Books may reflect impact in other conferences, which is a logical type of impact to assess.

Whilst many other types of altmetric have also been proposed, including a variety of social media metrics, the above four seem to be the most promising and form a relatively coherent conference-related set to explore in a first investigation of conference-related altmetrics.

3.3 Mendeley readers

As reviewed above, many studies have investigated Mendeley reader counts as an altmetric for journal articles. They have usually followed a correlation approach and have found statistically significant moderate or strong positive correlations between Mendeley readership counts and citations counts for journal articles. Some investigations of Mendeley readership counts have also investigated books, finding moderate or weak correlations, but no previous study has investigated the relationship between Mendeley readership counts and citation counts for conference papers. This chapter partly fills this gap by investigating the correlation between Mendeley readership and Scopus citation counts for conference papers in four engineering fields in which conferences are important (Computer Science Applications; Computer Software; Building & Construction Engineering; and Industrial & Manufacturing Engineering). The chapter also investigates reasons for highly cited papers in Scopus with few Mendeley readers and papers with many Mendeley readers but few citations. This leads to the first research question.

- RQ1: Do Mendeley readership counts reflect the scholarly impact of conference papers in conference-based fields?

Since most prior altmetrics research has focused on journal articles, it is also useful to explicitly compare the conference-based results with equivalent journal-based results. This

could reveal whether conferences tend to be ignored in Mendeley, for example, or whether conference-based *fields* tend to ignore Mendeley. This is important because prior book-based studies of Mendeley have found lower coverage of books than journal articles and so seems that not all types of scholarly output are equally covered by Mendeley.

- RQ2: Does the answer to the above research question differ between fields in comparison to journal articles?

Although correlation tests are important first steps for analysing altmetrics, investigating outliers can reveal important differences between altmetrics and the baseline impact indicator, citation counts. Thus, this is a logical additional type of investigation.

- RQ3: What are the causes of conference papers having many Mendeley readers compared to citations or many citations compared to Mendeley readers?

3.4 Google Patents

Although Google Patents has been shown to be a useful source of patent citations to journal articles in some fields, including engineering-related areas (especially bioengineering, which is not covered here), it is not known whether there are patent citations to conference papers, or whether it is possible to extract them automatically. Prior research has developed heuristics to extract citation to journal articles but these may not work for conference papers. Thus, the first two patent-related questions relate to whether it is possible to find patent citations to conference papers and whether there are enough of them.

- RQ4: Can citations from Google Patents to conference papers be automatically extracted using curated Bing queries?

- RQ5: Are there enough citations from Google Patents to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Google Patents in conference-based fields?

As for Mendeley, it is important to correlate Google Patents citations with Scopus citations to assess the extent to which the two reflect similar types of impact and to validate Google Patent citations as non-random and scholarly-related. The following research question also covers disciplinary differences.

- RQ6: Do citations to conference papers from Google Patents reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Google Patents in conference-based fields?

There is no equivalent to RQ3 (causes of outliers) for Google Patents because Google Patents citations are too rare to cause statistical outliers.

3.5 Wikipedia citations

No studies have investigated Wikipedia citations to conference papers and this is investigated here. Similarly to the case for patent citations, the heuristics used to extract Wikipedia

citations need to be shown to work, it is important to assess the results in comparison to Scopus citations and journal articles.

- RQ7: Can Wikipedia citations to conference papers be automatically extracted using curated Bing queries?
- RQ8: Are there enough citations from Wikipedia to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Wikipedia in conference-based fields?
- RQ9: Do citations to conference papers from Wikipedia reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Wikipedia in conference-based fields?

3.6 Google Books citations

The fourth gap in the literature is Google Books citations to conference papers. Its API can be used to automatically search for citations, but manual checking is needed to filter out false matches (Kousha & Thelwall, 2015a) and it is not known whether these heuristics will work for conference papers. Moreover, it is also not known if conference papers tend to be extensively cited by books. As for patent and Wikipedia citations, the heuristics used to extract Google Books citations need to be shown to work, it is important to assess the results in comparison to Scopus citations and journal articles.

- RQ10: Can citations from Google Books to conference papers be automatically extracted using Google Books API queries and heuristics to filter the results?
- RQ11: Are there enough citations from Google Books to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Google Books in conference-based fields?
- RQ12: Do citations to conference papers from Google Books reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Google Books in conference-based fields?

3.7 Summary

Bearing in mind the limitations of the previous study and the gaps in the literature, this thesis investigates indicators for the impact of conference papers and journal articles using four different types of web indicators: Mendeley readers, Google Patents citations, Wikipedia citations and Google Books citations. The thesis analyses journal articles and conference papers in four selected engineering fields (Computer Science Applications, Software Engineering, Building & Construction Engineering and Industrial & Manufacturing Engineering), comparing the web indicators with Scopus citation counts. This thesis also investigates why articles in Mendeley are widely read but have few Scopus-indexed citations, or highly cited but with few Mendeley readers, in conference papers. The thesis also assesses

the accuracy of methods to count citations from Google Patents, Wikipedia and Google Books to conference papers.

In summary, the primary goals of the study as captured from the research questions is to assess the value of four promising altmetrics for conference papers in four engineering-related fields in which conferences seem to be important.

Chapter 4: Research Methods

4.1 Introduction

The research design for the thesis was to obtain large sets of conference papers and articles from the four chosen fields and to calculate and compare their Scopus citation counts with their altmetric scores. Narrow subject categories were used to ensure more comprehensive coverage. From the Scopus computer science category, a field in which conferences are arguably more important than journals, the two categories of Computer Science Applications and Computer Software Engineering were chosen. Conferences are also known to be important in engineering in general, and so from the broad Scopus Engineering category, the two narrow categories of Building & Construction Engineering and Industrial & Manufacturing Engineering were chosen.

Articles and conference papers were downloaded from Scopus for the year 2011. This year was chosen to give a long enough period to attract a substantial number of citations and altmetric scores from the potentially slower sources (Google Patents, Google Books) so that a lack of results could not be attributed to the conference papers being too young. The same fields and year were investigated for all altmetrics so that the results could be compared between them.

4.2 Scopus data collection

Bibliographic information and citation counts the most recent 5000 journal articles, conference papers and the oldest 5000 journal articles and conference papers from each field (Computer Science Applications; Computer Software; Building & Construction Engineering; Industrial & Manufacturing Engineering) from 2011 were downloaded from Scopus during March 2015.

The Scopus API allowed a maximum of 5000 results to be downloaded for each query (in 2018, after the studies in this thesis had been completed, this restriction was relaxed). These 5000 can either be the first 5000 or last 5000 in the year. If there are up to 10000 query matches then a complete set can be downloaded by submitting two queries, one for the first 5000 and one for the last 5000, and combining the results. This formed a complete set for Building & Construction Engineering.

For three fields (Computer Science Applications; Computer Software; Industrial & Manufacturing Engineering) there were more than 10000 query matches. In these three cases, the two queries resulted in a time balanced set of 10000 articles, the first and last 5000 of the year. Although this is an incomplete set of data, this should not affect the correlation results, citation averages or the proportions cited because the set is time balanced. Thus, articles in the middle of the year should have properties that are the average of the early and late articles in the year and the overall correlation should be approximately the same as if the middle articles were also included. This is based on the assumption that the middle articles are not unusual in some way. They could be unusual, for example, if low quality conferences tended to occur in the middle, start or end of the year. This restriction to 10,000 strategy has the

limitation that conferences in the middle of a year for large categories may be omitted, and these may be particularly prestigious in some fields.

The same Scopus data set was used for comparison with Mendeley readers, Google Patents citations, Wikipedia citations and Google Books citations so that the results would be comparable between altmetrics. All altmetric scores were obtained using the metadata of the publications found in the four Scopus datasets.

4.3 Mendeley data collection

All articles and conference papers were submitted to the Mendeley API via Webometric Analyst to count the number of users in Mendeley that had registered articles and conference papers in the four Scopus engineering fields from the year 2011. Conference papers and journal articles were found in Mendeley with a search for the publication year, the first author last name, and the title, as in the following example. *title: "Experimental verification of hysteresis in gait transition of a quadruped robot driven by nonlinear oscillators with phase resetting AND author: Shinya AND year: 2011".* The year 2011 was chosen to give a substantial period for citations to accrue so that there is more chance of getting high correlations between citation and readership counts for both journal articles and conference papers.

The Mendeley queries could produce incorrect matches but these were removed with the following automatic checks in Webometric Analyst:

1. Articles with Digital Object Identifier (DOI) for both Scopus and Mendeley were compared and if found identical, then those articles were classified as correct matches. But if found to be not identical, then those articles are classed as incorrect matches. Thus, before comparison, both DOIs from Scopus and Mendeley were all converted to lower case, removed all spaces, and any DOI initial or http://dx.doi.org/ was removed.
2. Checking for the publication years, and if found to differed between two sites the article was classified as false match. This step is applicable to articles that have no DOI for both Scopus and Mendeley.
3. The third step compared the first author last name in the Mendeley record and the first author last name in Scopus, after converting both to lower case, and removing any spaces, hyphens and accent marks. Thus, the article will be rejected if the first author last name in Mendeley was not a substring of the first author last name in Scopus. This partial matching was used instead of the exact matching for countries that use double last names and Mendeley users may use only one of the names.
4. Titles words were compared, after removing accents from the letter, leaving all the punctuation. The articles were rejected if more than 15% of the words in one version of the title could not be found in the other version of the title and vice versa (the average of the two calculations). For articles with a dual title in Scopus (e.g., Spanish and English variants) separated by a | symbol, the titles were matched three times: once with the English version, once with the other version, and once with both together. The article was kept if any one of the three comparisons resulted in at least 85% commonality of the

words in the title. This gives a match if the Mendeley user enters either language version of the title or both (Thelwall & Wilson, 2016).

4.4 Google Patents citations data collection

Bing searches were used to extract and filter Google Patents citations in December 2016, since Google does not allow large-scale direct automatic patent searching. The free computer program Webometric Analyst (http://lexiurl.wlv.ac.uk) was used to perform automatic searches with the Bing API by searching for key bibliographic information from a large set of patents. From the "Make Searches" menu the option "Make Google Patent Searches for a set of Scopus/WOS/ other journal articles or books" was used. This generates queries with the last names of the authors (up to a maximum of eight) and the conference paper or journal article title as a phrase search along with the publication year and the command *site:google.com/patents/* to restrict the results to patents in the Google Patents site.

Ruiz "DoubleFlip A motion gesture delimiter for mobile interaction" 2011 site:google.com/patents/ ***0***

Knievel Noemm Hoeher "Low-complexity receiver for large-MIMO space-time coded systems" 2011 site:google.com/patents/ 1

Hinckley Song "Sensor synaesthesia Touch in motion and motion in" 2011 site:google.com/patents/ ***0***

Kramer Majidi Sahai "Soft curvature sensors for joint angle proprioception" 2011 site:google.com/patents/ ***1***

The patent citation queries described above were searched in Bing using Webometric Analyst and the number of matches recorded for each one.

4.5 Wikipedia citations data collection

For Wikipedia citations, we re-used the Scopus 2011 data for the four engineering fields (Computer Science Applications; Computer Software; Building & Construction Engineering; and Industrial & Manufacturing Engineering).

Bing searches were used to extract and filter Wikipedia citations in December 2016, since Wikipedia does not allow large-scale direct automatic citation searching. The free Webometric Analyst software (http://lexiurl.wlv.ac.uk) was used to perform automatic searches with the Bing API by searching for the key bibliographic information of a large set of articles and papers. From the "Make Searches" menu the option "Make Wikipedia Searches for a set of Scopus/WoS/ other journal articles or books" was used. This generates queries with the last names of the authors (up to a maximum of eight) and the conference paper or journal article title as a phrase search along with the publication year and the command site:wikipedia.org/wiki/ to restrict the results to the Wikipedia site, as the following examples illustrate.

Alessandrini Holguin Parent "Advanced transport systems showcased in La Rochelle" 2011 site:wikipedia.org/wiki/ ***1***

Gálvez-López "Real-time loop detection with bags of binary words" 2011 site:wikipedia.org/wiki/ ***0***

Bautista-Gomez Komatitsch Maruyama "FTI High performance fault tolerance interface for hybrid" 2011 site:wikipedia.org/wiki/ ***1***

Watanabe Kanou Kobayashi "Development of a steerable drill for ACL reconstruction" 2011 site:wikipedia.org/wiki/ ***1***

The above queries were searched in Bing using Webometric Analyst and all the matches were recorded for each one.

4.6 Google Books citations data collection

The Google Books API was used to extract and filter Google Books citations in March 2017 (see Kousha & Thelwall, 2009). Webometric Analyst (http://lexiurl.wlv.ac.uk) was used to perform automatic searches with the Google Books API by searching for key bibliographic information. For example, this conference paper by Das Maughan McCann "Towards mixed-initiative multi-robot field experiments Design deployment and lessons learned" 2011 have been cited by three books in Google Books:

https://books.google.com/books?isbn=1461456592
https://books.google.com/books?isbn=3319237780
https://books.google.com/books?isbn=3319000659.

The following process was used:

Step One: From the 'Books' tab menu the option 'make Google Books queries from Scopus, WoS or other Data' was used. The queries are manually checked for data from Scopus or WoS file, authors name, Book title, publication year and publication title were all checked, and the query is then saved and submitted for the next step.

Step Two: 'Search Google Books with all queries in file' is where the computer program Webometric Analyst will automatically generate the queries.

Step Three: 'Remove GBS matches query terms', this task removed search matches that did not mention the correct books.

Step Four: 'Add filtered GBS matches to Original Result file', this step filters out any false matches to the original file.

Step Five: 'Add citation to GBS summary'. This is the last step that summarises the submitted query and the number of citations extracted.

4.7 Altmetric accuracy checks

As described above, accuracy checks are built in to Webometric Analyst for Mendeley readers and so were only carried out for the other three altmetrics.

Google Books: Data checks were done on random samples of results extracted from each subject category using Webometric Analyst. Google Books citations were randomly selected using Webometric Analyst to select 25 samples from each subject category. Every query was tested in Google Books to check if the book returned by the API had a genuine citation to the conference paper. This manual check was performed by matching the text of the citation to the known information about the conference paper or journal article (e.g., title, publication year, authors).

Wikipedia citations: All Wikipedia citations were checked manually by visiting the source Wikipedia page and searching it for the (apparently) referenced conference paper or journal article. If no matches had been found then the edit history would have been consulted for previous versions, but this was not necessary.

Google Patent citations: All Google Patent citations were checked manually by visiting the source Google Patent page and searching it for the (apparently) referenced conference paper or journal article.

4.8 Averages and proportions cited

For all altmetrics and Scopus citations, averages and proportions cited were calculated for conference papers and journal articles as basic descriptive information. A geometric mean was calculated in preference to an arithmetic mean for average altmetric scores and Scopus citations, because the data (particularly Mendeley readers and Scopus citations) is skewed. It was calculated by taking the logarithmic transformation $ln(1+x)$ on the data set for both patent citations and citation counts, and then the average of the transformed data was taking using the formula $Exp(Average)-1$ to determine the geometric mean. Here *Average* is the arithmetic mean of the log-transformed citation counts.

4.9 Correlation tests

Spearman correlations were used to compare Scopus citations and altmetric scores because citation data are too skewed for the normality assumption of a Pearson test. The Spearman rank correlation is a non-parametric test that is used to measure the degree of association between two variables. The Spearman rank correlation test does not carry any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal. Spearman correlations were calculated in the SPSS software.

Rank correlation measures the strength and direction of rank association between two variables on a scale +1 and -1. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker. The direction of the relationship is indicated by the sign of the coefficient: a + sign indicates a positive relationship and a – sign

indicates a negative relationship. Correlations between an altmetric score and citation counts are the most practical technique that help to validate a research indicator (Lee, Sugimoto & Zhang, 2013; Sud & Thelwall, 2014). Correlation tests have a long tradition in webometrics to evaluate the evidence provided by web data for individual articles or journal web sites (Vaughan & Hysen, 2002; Vaughan & Shaw, 2005). Correlation tests are also standard in altmetrics, playing a similar role for research validation (Li, Thelwall & Guistini, 2012).

A positive correlation with citation counts gives evidence that the indicator at least reflects academic impact to some extent and probably also academic quality. This is because citation counts are known to partly reflect academic quality to some extent in most fields and so should correlate positively with any other indicator that also correlates with research quality.

To calculate 95% confidence intervals for the correlation coefficients, the standard formula was used: $\tanh(\operatorname{arctanh}(r) \pm \frac{1.96}{\sqrt{n-3}})$. Here, r is the sample correlation and n is the sample size. (See appendices 16-22). A 95% CI was found by adding and subtracting 1.96 times its standard error (Dowdy & Chilko 2011).

4.10 Mendeley outlier analysis

To identify Mendeley outliers (articles with many Mendeley readers and few Scopus citations or vice versa), linear regression was used to regress reader counts against citation counts for both articles and conference papers in each subject area, this step is to estimate the expected number of Mendeley readers for each conference paper based upon its citation count (Thelwall & Wilson, 2014a).

The logarithmic transformation $ln(1 + x)$ was used in the data set for both readers and citation counts to reduce the skewness of the data, before regressing the reader counts against the citation counts (Thelwall & Wilson, 2014b) to get a statistically robust regression fit. The purpose of the logarithmic transformation was to avoid focusing too much on papers with high values of one or other variable but to include outliers where the anomalies were within the low or moderate end of the scale. Thus, for example, a difference between 0 and 4 citations seems more significant than a difference between 40 and 44 citations, whereas the differences would be the same for untransformed data.

Any transformation affects the nature of the outliers to be examined. Untransformed data (i.e., a conference paper would be an outlier based on a straightforward comparison of its numbers of readers and citations) would not be helpful since it would not be possible to get an accurate regression equation for the highly skewed data.

Although there are different methods to detect outliers in a data set, the residual method seems to be the simplest and straightforward used in this thesis. The Residual method is simply "the difference between the number of readers of an article and the number of readers estimated by its number of citations using regression" (Thelwall & Wilson, 2016). Thus, this is a reasonable method to detect outliers. Its main limitation is that it is biased towards large numbers in the sense that the residuals of the same size are probably inherently more important for articles with fewer citations because the ratio of the number of readers to

citations would be higher. The Studentised variant of the residual was used, which calculates residuals after regressing against all data except the point considered.

The residuals from the linear regression were taken as the main outliers. Since Mendeley readers are anonymous in the data set it would not have been possible to contact any to ask them why they were interested in an article or conference paper. Thus, the outlier analysis was based on the properties of the cited documents.

The outliers were qualitatively investigated for likely causes by repeatedly reading the titles and abstracts of all the outlier papers and looking for possible reasons why they had attracted relatively many or few Mendeley readers. This qualitative investigation was informed by similar prior studies but is error prone and may miss important patterns.

4.11 Summary

In summary, data sets of up to 10000 articles and conference papers from 2011 were extracted from Scopus for four fields and corresponding altmetric scores obtained for Mendeley, Google Patents, Wikipedia and Google Books using Webometric Analyst Bing searches or API queries. The results were then manually checked (Google Patents, Wikipedia and Google Books). Averages and proportions cited were also calculated to compare between fields and between journal articles and conference papers.

Correlation tests were used to assess the non-randomness and scholarly-relatedness of the altmetrics in all fields. Outlier analyses were used for the altmetric with high scores (Mendeley) to get insights into whether Mendeley may reflect a different type of impact to that of Scopus.

Chapter 5: Results

This chapter reports the main thesis results and an initial analysis. Deeper analyses are reported in the Discussion chapter.

5.1 Mendeley readership counts for conference papers (RQ1, RQ2)

The correlation tests found that the Mendeley readership counts correlate strongly (0.560-0.662) with citation counts in all subject categories for journal articles (Table 5.1). For conference papers, readership counts correlate moderately (0.437-0.439) with citation counts in Computer Science Applications and Software. Readership counts have low correlations (0.143-0.168) with citation counts in Building & Construction and Industrial & Manufacturing Engineering. The low correlations for conference papers in Building & Construction Engineering (0.143) and Industrial & Manufacturing Engineering (0.168) might be due to the low coverage of conference proceedings in engineering subject categories, reducing their Scopus citation counts. Correlation tests were used between Mendeley readers and citation counts to determine the relationship between readership counts and citation counts for both conference papers and journal articles. Significant positive correlations suggest that there may be a common factor between readership and citation counts. Proportions test were also calculated on each subject category for both journal articles and conference papers to determine the percentage of papers with Scopus citations and Mendeley readers.

Table 5.1. Spearman correlations between Mendeley reader counts and citation counts for articles and conference papers in Scopus from 2011 in the four subject categories analysed. 95% confidence intervals are reported underneath each correlation. Categories with 9999 or 10000 articles are incomplete (the first and last 5000 articles/papers in the year) whereas the remaining categories are complete.

Scopus subject category	Articles	Conference papers	Spearman correlation for articles (95% CI)	Spearman correlation for papers (95% CI)
Computer Science Applications	10000	9999	0.560** (0.546,0.573)	0.439** (0.423,0.455)
Computer Software	10000	9974	0.572** (0.559,0.585)	0.437** (0.421,0.453)
Building & Construction Eng.	8433	4750	0.662** (0.650,0.674)	0.143** (0.115,0.171)
Industrial & Manufacturing Eng.	10000	9999	0.660** (0.649,0.671)	0.168** (0.149,0.187)

There are relatively low proportions of cited papers in Scopus for Industrial & Manufacturing Engineering (17.5%) and Building & Construction Engineering (18.3%) conference papers (Table 5.2). This could be due to low coverage in Scopus of conference proceedings. Also, there are low Mendeley reader counts for Building & Construction Engineering (18.7%)

conference papers, which could be due to few Mendeley users or the low value of conference papers in this field.

Table 5.2. Scopus citation counts and Mendeley readership counts, median, geometric mean and percentage coverage for both journal articles and conference papers.

Scopus subject category	**Journal articles**		**Conference papers**	
	Scopus median, geometric mean % cited	*Mendeley* median geometric mean % with readers	*Scopus* median geometric mean % cited	*Mendeley* median, geometric mean % with readers
Computer Science Applications	3 0.91 80.9%	3 0.93 64.8%	0 1.20 34.4%	0 0.41 47.2%
Computer Software	3 0.91 80.6%	3 0.88 62.1%	3 1.43 54.7%	10 1.02 68.6%
Building & Construction Eng.	2 0.71 71.7%	2 0.65 52.7%	0 0.08 18.3%	0 0.09 18.7%
Industrial & Manufacturing Eng.	2 0.73 71.3%	2 0.64 55.9%	0 0.08 17.5%	0 0.28 41.0%

5.2 Outliers between Mendeley readers and Scopus citations (RQ3)

Outliers between Mendeley readers and Scopus citation counts were checked using the logarithmic transformation $ln(1 + x)$ in the data set for both readers and citation counts to reduce the skewness of the data, before regressing the reader counts against the citation counts. Then the residuals from the linear regression were used to check the main outliers. These were then manually investigated qualitatively for likely causes (Table 5.3, Table 5.4).

The following were identified in the study as likely causes of high Mendeley readership counts compared to Scopus citation counts for conference papers.

a) Papers that are written based on improving the performance of an existing system. Computer Science Applications; *"Purlieus: Locality-aware resource allocation for MapReduce in a cloud*", has 74 Mendeley readers but no citations. *"Reducing electricity cost through Virtual Machine placement in high performance computing clouds*", (66 readers, 0 citations) demonstrates a system that can be used to reduce electricity cost and load migration at minimum low electricity consumption rates. " *On the duality of data-Intensive file system design: Reconciling HDFS and PVFS"* (62 readers, 0 citations).
b) Papers that create public awareness, motivation and participation for new scientific discoveries. For example, three papers from Computer Software; *"Attention please! Learning analytics for Visualization and recommendation"*, (202 readers, 1 citation) *"Dusting*

*for Science: Motivation and participation of digital citizen science volunteer*s", (149 readers, 6 citations) and *"A survey of risks, threats and vulnerabilities in cloud computing"*, (115 readers, 0 citations).

c) Papers that are relevant to daily life or religious beliefs. For example, paper in Industrial & Manufacturing Engineering, *"Halal supply chains in the food industry- A conceptual model"*, (50 readers, 6 citations).

d) Practical solutions to important real-world problems. For example, Building & Construction Engineering, *"Overview of UFC 3-340-02, Structures to resist the effects of accidental explosions"*, has 29 Mendeley readers but no citation. *"Sandnet: Network traffic analysis of malicious software"* (36 readers, 10 citations) In the same Building & Construction Engineering Scopus subject category for conference paper, *"Exploiting home automation protocols for load monitoring in smart buildings"* (32 Mendeley readers 1 citation).

e) Social media articles that may be of general interest to users. For example, in Computer Science Applications article papers; *"Serious social media: On the use of social media for improving student's adjustment to college*" (170 readers, 15 citations), and *"Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning"* (404 readers, 74 citations).

f) Practical commercial advice. In Building & Construction Engineering, an article titled; *"Characterizing entry mode for international construction markets: paving way to a selection model"* (16 readers, 0 citations).

g) Articles of regional interest. For example, Computer Software, *"A citizen-oriented approach for evaluating the performance of e-government in Sri Lank*a" (28 readers, 0 citations).

The following were identified as likely causes of high Scopus citation counts compared to Mendeley readership counts for conference papers

h) Papers on software packages that may be cited if the software is used, without necessarily reading the paper. In Computer Software, *"MICE: Multivariate Imputation by Chained Equation in R"* (249 citations, 0 readers) and "*ContextFJ: A minimal core calculus for context-oriented programming*" (18 citations, 0 readers)

i) Papers with a set model for completing a task. For example, in Computer Science Applications, *"Recommended steps for thematic synthesis in software engineering"* (20 citations, 0 readers). *"A framework for capturing distinguishing user behaviours in novel interfaces"* (13 citations, 0 readers) and Building and Construction Engineering *"A naming convention for the piano key weirs geometrical parameters"* (20 citations, 0 readers).

Table 5.3. Outliers between Mendeley readers and Scopus citation counts for conference papers in four Scopus subject categories.

Field	Read	Cited	Resid.	Paper title
Computer Science Applications	62	0	1.59	On the duality of data-intensive file system design Reconciling HDFS and PVFS
	74	0	1.67	Purlieus: Locality-aware resource allocation for MapReduce in a cloud
	66	0	1.62	Reducing electricity cost through Virtual Machine placement in high performance computing clouds
	0	11	-1.41	Understanding net zero energy buildings: Evaluation of load matching and grid interaction indicators
	0	13	-1.09	A framework for capturing distinguishing user behaviours in novel interfaces.
	0	16	-1.16	Recommended steps for thematic synthesis in software engineering
Software Engineering	202	1	1.64	Attention Please! Learning analytics for Visualization and recommendation
	149	6	1.14	Dusting for Science: Motivation and participation of digital citizen science volunteers.
	115	0	1.61	A survey of risks, threats and vulnerabilities in cloud computing
	0	36	-1.53	CloudNet: Dynamic pooling of cloud resources by live WAN migration of virtual machines
	0	18	-1.33	ContextFJ: A minimal core calculus for context-oriented programming
	0	26	-1.43	The development and testing of human machine interface for a mobile medical exoskeleton
Building & Construction Engineering	29	0	1.41	Overview of UFC 3-340-02, structures to resist the effects of accidental explosions
	36	10	1.29	Sandnet: Network traffic analysis of malicious software
	32	1	1.39	Exploiting home automation protocols for load monitoring in smart buildings
	0	18	-0.33	A naming convention for the piano key weirs geometrical parameters
	0	20	-0.34	An analytical method to calculate borehole fluid temperatures for time-scales from minutes to decades.
	0	18	-0.33	The redesign of Italian building to reach net zero energy performances: A case study of the SHC Task4- ECBCS Annex52.
Industrial & Manufacturing Eng.	50	6	1.66	Halal supply chain in the food industry A conceptual model
	44	0	1.43	Understanding of defect physics in polycrystalline photovoltaic materials
	61	3	1.34	Key performance indicators for sustainable manufacturing evaluation in automotive companies
	0	14	-0.66	The effects of stock index futures to stock market volatility
	0	10	-0.61	Combined electric light and daylight systems ecodesign
	0	9	-0.60	A review of engineering research in sustainable manufacturing

The following were identified in the study as likely causes of high Mendeley readership counts compared to Scopus citation counts for journal articles. These reasons are very tentative since they are based on a single paper.

a) Papers on systems for context awareness. In Software Engineering**, "***Context-aware recommender systems"* (385 readers, 17 citations).
b) Papers on perfect modern accommodations. In Building & Construction Engineering, "*Three-dimensional study for evaluating of air flow movements and thermal comfort in a model room: Experimental validation"* (25 readers, 0 citations)
c) Papers on systems for studying wildlife. In Computer Science Applications, *"CARNIVORE: A disruption-tolerant system for studying wildlife"* (51 readers, 0 citations)

The following were identified as likely causes of high Scopus citation counts compared to Mendeley readership counts for journal articles. These reasons are also very tentative since they are based on a single paper.

d) Papers on computerised method for solving difficult mathematical problems. In Computer Science Applications, *"Iterative methods for solving nonconvex equilibrium variational inequalities"* (30 citations, 0 readers)
e) Papers on cloud manufacturing and services. In Industrial & Manufacturing Engineering, *"Typical characteristics of cloud manufacturing and several key issues of cloud service composition"* (citations 74, 0 readers)

Table 5.4. Outliers between Mendeley readers and Scopus citation counts for journal articles in four Scopus subject categories.

Field	Read	Cited	Resid	Paper title
Computer Science Applications	170	15	1.21	Serious social media: On the use of social media for improving students' adjustment to college.
	51	0	1.48	CARNIVORE: A disruption-tolerant system for studying wildlife.
	404	74	1.15	Personal learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning
	0	30	-1.21	Iterative methods for solving nonconvex equilibrium variational inequalities.
	0	48	-1.34	Contact treatment in isogeometric analysis with NURBS.
	0	25	-1.16	Replenishment run time problem with machine breakdown and failure rework
Software Engineering	28	0	1.27	A citizen-oriented approach for evaluating the performance of e-government in Sri Lanka.
	385	17	1.54	Context-aware recommender systems
	55	0	1.56	Engineering privacy revisited
	0	25	-1.16	Distributed event-triggered tracking control of leader-follower multi-agent systems with communication delays
	0	249	-1.83	mice: Multivariate imputation by chained equations in R
	0	20	-1.09	An interval set model for learning rules from incomplete information table
Building & Construction Engineering	34	2	1.09	Adaptive reuse of heritage buildings
	25	0	1.33	Three-dimensional study for evaluating of air flow movements and thermal comfort in a model room: Experimental validation
	16	0	1.15	Characterizing entry mode for international construction markets: paving way to a selection model
	0	38	-1.32	A new method for optimal selection of sensor location on a high-rise building using simplified finite element mode
	2	66	-1.02	Optimal sensor placement for structural health monitoring based on multiple optimization strategies
	0	15	-1.02	Mechanical properties of self-company concrete incorporating quarry dust powder, silica fume or fly ash
Industrial & Manufacturing Engineering	50	3	1.18	The determinants of merger waves: An international perspective
	34	0	1.43	The Item Response Theory: Possible contributions to marketing studies
	24	0	1.28	Newsvendor pricing problem in a two-sided market
	0	27	-1.12	Recent development in finite element analysis of self-piercing riveted joints
	0	74	-1.41	Typical characteristics of cloud manufacturing and several key issues of cloud service composition
	0	33	-1.17	Migration of mineral oil from printed paperboard into dry foods: Survey of the German market

Table 5.3 shows readership and citation counts for conference papers for the selected subject categories. There are more Mendeley readers than citations in all four subject areas. In contrast, some fields had more Scopus citations than Mendeley readers for journal articles (Table 5.4). Thus, in the Engineering fields, either Mendeley users are more likely to add journal articles than conference papers that they cite, or (as seems more likely) documents citing engineering conference papers are less likely to be indexed in Scopus than documents citing engineering journal articles. This seems more likely because conference papers are presumably more likely to cite other conference papers and are less well covered by Scopus than journal articles.

5.3 Google Patents citations data checks (RQ4)

All Google Patent citations found in all searches were checked and found to be correct (Table 5.5). Tables 5.6 to 5.9 confirm the patent citation checks by including a copy the citation as shown in the patent to validate its accuracy.

There were no false matches and, although the set checked is small, the automatic method to extract conference citations from Google patents seems to be very accurate. It is not known whether it finds all relevant conference citations, however (i.e., recall). It seems likely that it will miss some conference citations that are in unusual formats.

Table 5.5. The results of manual checks of the Bing search matches for Google patents citations to conference papers.

Scopus subject category	Citations	Correct citations
Computer Science Applications	53	53 (100%)
Software Engineering	25	25 (100%)
Industrial & Manufacturing Engineering	7	7 (100%)
Building & Construction Engineering	0	0 (-)

Table 5.6. Manual data checks of Google patents citations for Industrial & Manufacturing Engineering conference papers. All citations were valid.

Patent	Query	Citation in patent
Patent US20120226390 - History timeline display for ...	Pham "A framework algorithm for a real-world variant of" 2011 site:google.com/patents/	A framework algorithm for a real-world variant of the vehicle routing problem; Vu Pham; Tien Dinh; Industrial Engineering and Engineering Management (IEEM), 2011 IEEE International Conference on; Digital Object Identifier: 10.1109/IEEM.2011.6118237 Publication Year: 2011, Page(s): 1859 – 1863
Patent US8781004 - System and method for encoding video ...	Oh Lee Kim "An adaptive sharpening filter using quantization step size" 2011 site:google.com/patents/	Sye-Hoon Oh, et al. "An Adaptive Sharpening Filter Using Quantization Step Size and Pixel Variance in H.264/AVC", Consumer Electronics (ICCE), IEEE International Conference on Jan. 9, 2011.
Patent US9131073 - Motion estimation aided noise reduction ...	Oh Lee Kim "An adaptive sharpening filter using quantization step size" 2011 site:google.com/patents/	Sye-Hoon Oh, et al. "An Adaptive Sharpening Filter Using Quantization Step Size and Pixel Variance in H.264/AVC", Consumer Electronics (ICCE), IEEE International Conference on Jan. 9, 2011.
Patent US9080501 - Engine combustion control via fuel ...	Splitter Hanson Kokjohn "Reactivity controlled compression ignition RCCI heavy-duty engine operation" 2011 site:google.com/patents/	Splitter, D.A., Hanson, R.M., Kokjohn, S.L., and Reitz, R.D., "Reactivity Controlled Compression Ignition (RCCI) Heavy-Duty Engine Operation at Mid-and High-Loads with Conventional and Alternative Fuels," SAE Paper 2011-01-0363, 2011.
Patent WO2014041326 A1 - Transmission system - Google Patents	Moyers Akehurst Parker "The application of the milner CVT as a" 2011 site:google.com/patents/	MOYERS, J.; AKEHURST, S.; PARKER, D.A.; SCHAAF, S.: 'The Application of the Milner CVT as a Novel Power Splitting Transmission for Hybrid Vehicles' SAE TECHNICAL PAPER, 2011-01-0890 2011,
Patent WO2014041326 A1 - Transmission system - Google Patents	Sovran "The impact of regenerative braking on the powertrain-delivered" 2011 site:google.com/patents/	SOVRAN, G.: 'The Impact of Regenerative Braking on the Powertrain-Delivered Energy Required for Vehicle Propulsion' SAE TECHNICAL PAPER, 2011-01-0891 2011,

Table 5.7. Manual data checks of Google patents citations for Computer Software conference papers.

Patent	Query	Citation in patent
Patent US8798840 - Adaptive mapping with spatial summaries ...	Strom "Occupancy grid rasterization in large environments for teams" 2011 site:google.com/patents/	Johannes Strom et al. "Occupancy Grid Rasterization in Large Environments for Teams of Robots," 2011 IEEE/RSJ Int'l Conf. on Int. Robots and Systems, Sep. 2011 pp. 4271-4276.
Patent US8781739 - Systems and methods for using magnetic ...	Moore "Magnetic localization for perching UAVs on powerlines" 2011 site:google.com/patents/	Moore et al., "Magnetic Localization for Perching UAVs on Powerlines", International Conference on Intelligent Robots and Systems, San Francisco, California, Sep. 25-30, 2011, 8 pages.
Patent US8879831 - Using high-level attributes to guide ...	Girshick Shotton Kohli "Efficient regression of general-activity human poses from depth" 2011 site:google.com/patents/	Girshick, R. et al. Efficient regression of general-activity human poses from depth images. In Proc. ICCV, 2011.
Patent US9324112 - Ranking authors in social media systems	Pal "Identifying topical authorities in microblogs" 2011 site:google.com/patents/	Pal, A. et al.; "Identifying Topical Authorities in Microblogs"; Proceedings of the 4th ACM International Conference on Web Search and Data Mining; 2011; pp. 45-54.
Patent US9430043 - Bioacoustic control system, method and ...	Hinckley Song "Sensor synaesthesia Touch in motion and motion in" 2011 site:google.com/patents/	Hinckley, Ken, and Hyunyoung Song, "Sensor synaesthesia: touch in motion, and motion in touch." Proceedings of the Sigchi Conference on Human Factors in Computing Systems. ACM, 2011.
Patent US9201520 - Motion and context sharing for pen ...	Tashman "LiquidText A flexible multitouch environment to support active" 2011 site:google.com/patents/	Tashman, et al., "LiquidText: A Flexible, Multitouch Environment to Support Active Reading", In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, May 7, 2011, 10 pages.

Table 5.8. Manual data checks of Google patents citations for Computer Science Applications conference papers.

Patent	Query	Citation in patent
Patent US8798840 - Adaptive mapping with spatial summaries ...	Strom "Occupancy grid rasterization in large environments for teams" 2011 site:google.com/patents/	Johannes Strom et al. "Occupancy Grid Rasterization in Large Environments for Teams of Robots," 2011 IEEE/RSJ Int'l Conf. on Int. Robots and Systems, Sep. 2011 pp. 4271-4276.
Patent US8781739 - Systems and methods for using magnetic ...	Moore "Magnetic localization for perching UAVs on powerlines" 2011 site:google.com/patents/	Moore et al., "Magnetic Localization for Perching UAVs on Powerlines", International Conference on Intelligent Robots and Systems, San Francisco, California, Sep. 25-30, 2011, 8 pages
Patent US9312929 - System and methods to compensate for ...	Knievel Noemm Hoeher "Low-complexity receiver for large-MIMO space-time coded systems" 2011 site:google.com/patents/	C. Knievel, M. Noemm, and P. A. Hoeher, Low Complexity Receiver for Large-MIMO Space Time Coded Systems, in Proc. IEEE VTC-Fall'2011, Sep. 2011.
Patent US8069393 - Method and system for providing long ...	Eroz "Scrambled coded multiple access" 2011 site:google.com/patents/	Mustafa Eroz, Method and system for providing long and short block length low density parity check (LDPC) codes
Patent US8514825 - System and method for enabling a ...	Ibars Milito Monclus "Radio resource allocation for a high capacity vehicular" 2011 site:google.com/patents/	Ibars, Christian et al., "Radio Resource Allocation for a High Capacity Vehicular Access Network," 4th International Symposium on Wireless Vehicular Communications: WIVEC2011, Sep. 5-6, 2011, San Francisco, CA; U.S., 5 pages, http://www.ieeevtc.org/wivec2011/
Patent US20150274062 - Vehicle monitoring system - Google ...	Johnson "Driving style recognition using a smartphone as a" 2011 site:google.com/patents/	D.A. Johnson & M.M. Trivedi, "Driving Style Recognition Using a Smartphone as a Sensor Platform", 14 Int'l IEEE Conf. on Intelligent Transportation Sys. 1609-1615 (Oct. 2011)

5.4 Google Patents citations for conference papers (RQ5, RQ6)

Patent citations are rare for both journal articles and conference papers in the four engineering subjects (Table 5.9). Nevertheless, they occur in similar numbers for conference papers and journal articles, despite journal articles attracting far more citations in all fields. Thus, the lower scholarly impact of engineering conference papers does not translate into lower applied impact, in the form of patent citations.

Table 5.9. Google patent citations geometric mean and percentage coverage for both journal articles and conference papers.

Scopus subject category	**Journal articles**		**Conference papers**	
	Google Patents citations geometric mean (% cited)	Scopus citations geometric mean (% cited)	Google Patents citations geometric mean (% cited)	Scopus citations geometric mean (% cited)
Computer Science Applications	0.00124 (0.17%)	3.46 (80.9%)	0.00325 (0.31%)	0.53 (34.4%)
Computer Software	0.00197 (0.26%)	3.43 (80.6%)	0.00182 (0.22%)	1.29 (54.7%)
Industrial & Manufacturing Engineering	0.00077 (0.10%)	2.54 (71.3%)	0.00088 (0.11%)	0.18 (17.5%)
Building & Construction Engineering	0.00125 (0.13%)	2.43 (71.7%)	0 (0%)	0.20 (18.3%)

Some prior research has attempted to validate patent citation counts, such as by comparing them with journal citation counts. Papers with many citations from journal articles are more likely to be cited by patents (Meyer *et al.*, 2010). As shown in Table 5.10, there are statistically significant correlations between Scopus citation counts and Google Patents citations only in Computer Science Applications. In the other three subject categories, the correlations are not statistically significantly different from zero. In Computer Science Applications, the correlation is higher for journal articles than for conference papers but in both cases the correlations are low. This is due to the low numbers of patent citations, which makes it difficult to obtain a high correlation (Thelwall, 2016b).

Table 5.10. Spearman correlations between Google Patents citations, and Scopus citation counts for articles and conference papers in Scopus from 2011 in four engineering subjects.

Scopus subject category	Articles	Conf. papers	Spearman correlation for conference papers (95% CI)	Spearman correlation for journal articles (95% CI)
Computer Science Applications	8150	6698	0.072** (0.048, 0.096)	0.210** (0.179, 0.221)
Software Engineering	8232	8083	0.018 (-0.004, 0.040)	0.015 (-0.007, 0.037)
Industrial & Manufacturing Engineering	7355	5651	-0.015 (-0.041, 0.011)	-0.006 (-0.029, 0.017)
Building & Construction Engineering	8407	1754	0.000 (-)	0.005 (-0.016, 0.0026)

**Statistically significant at p=0.01

The positive correlation between Google Patents citations and Scopus citation counts for Computer Science Applications journal articles (0.210), corroborates past results (Kousha & Thelwall, 2015c), where the Spearman correlations between Google Patents citation counts and Scopus citation counts for Computer Science articles published between 1996–2012 was found to be (0.233). This positive significant correlation between Google Patents citations and Scopus citations for Computer Science articles shows that patents citation might partly reflect scholarly impact in Computer Science.

5.5 Wikipedia citations data checks (RQ7)

All automatically extracted citations from Wikipedia to conference papers were found to be correct (Table 5.11; see Tables 5.12-5.15 for individual check results). Thus, there seem to be no problems with the automatic method for identifying Wikipedia citations to conference papers.

Table 5.11. The results of manual checks of the Bing search results for Wikipedia citations to conference papers.

Scopus subject category	Citations	Correct citations
Computer Science Applications	35	35 (100%)
Software Engineering	168	168 (100%)
Industrial & Manufacturing Engineering	9	9 (100%)
Building & Construction Engineering	7	7 (100%)

Table 5.12. Manual data checks of Wikipedia citations for Computer Science Applications (conference papers).

Wiki page	Query	Citation in Wikipedia
Routing Protocol for Low power and Lossy Networks – Wikipedia	Clausen Herberg Philipp "A critical evaluation of the IPv6 Routing Protocol" 2011 site:wikipedia.org/wiki/	Clausen, T.; Herberg, U.; Philipp, M.; "A critical evaluation of the IPv6 Routing Protocol for Low Power and Lossy Networks (RPL)", Wireless and Mobile Computing, Networking and Communications (WiMob), 2011 IEEE 7th International Conference on , vol., no., pp.365-372,
Game theory - Wikipedia	Dolev Panagopoulou Rabie "Rationality authority for provable rational behavior" 2011 site:wikipedia.org/wiki/	Dolev, Shlomi; Panagopoulou, Panagiota; Rabie, Mikael; Schiller, Elad Michael; Spirakis, Paul (2011), "Rationality authority for provable rational behavior", Acm Podc: 289–290
Software analytics - Wikipedia	Hullett Nagappan Schuh "Data analytics for game development NIER track" 2011 site:wikipedia.org/wiki/	Kenneth Hullett, Nachiappan Nagappan, Eric Schuh, and John Hopson, "Data Analytics for Game Development (NIER Track)". In Proceedings of the International Conference on Software Engineering, May 2011, pp. 940-943.
Distributed file system for cloud - Wikipedia	Kobayashi Mikami Kimura "The gfarm file system on compute clouds" 2011 site:wikipedia.org/wiki/	Kobayashi, K; Mikami, S; Kimura, H; Tatebe, O (2011). The Gfarm File System on Compute Clouds. Parallel and Distributed Processing Workshops and PhD Forum (IPDPSW), 2011 IEEE International Symposium on. Grad. Sch. of Syst. & Inf. Eng., Univ. of Tsukuba, Tsukuba, Japan
Routing Protocol for Low power and Lossy Networks ...	Clausen Herberg Philipp "A critical evaluation of the IPv6 Routing Protocol" 2011 site:wikipedia.org/wiki/	T. Clausen, U. Herberg e M. Philipp, "A Critical Evaluation of the IPv6 Routing Protocol for Low Power and Lossy Networks (RPL)"
Resilient control systems - Wikipedia	Lin Sedigh Hurson "An agent-based approach to reconciling data heterogeneity in" 2011 site:wikipedia.org/wiki/	Lin, J.; Sedigh, S.; Hurson, A.R. (May 2011), An Agent-Based Approach to Reconciling Data Heterogeneity in Cyber-Physical Systems, 25th IEEE International Symposium on
Vector processor - Wikipedia	Kunzman "Programming heterogeneous systems" "IEEE International Symposium on" 2011 site:wikipedia.org/wiki/	Kunzman, D. M.; Kale, L. V. (2011). "Programming Heterogeneous Systems". 2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and PhD Forum. p. 2061
Distributed file system for cloud - Wikipedia	Undheim Chilwan Heegaard "Differentiated availability in cloud computing SLAs" 2011 site:wikipedia.org/wiki/	A., Undheim; A., Chilwan; P., Heegaard (2011). "Differentiated Availability in Cloud Computing SLAs". 2011 IEEE/ACM 12th International Conference on Grid Computing. pp. 129–136

Table 5.13. Manual data checks of Wikipedia citations for Computer Software (conference papers).

Wiki page	Query	Citation in Wikipedia
Masakatsu G. Fujie - Wikipedia	Watanabe Kanou Kobayashi "Development of a steerable drill for ACL reconstruction" 2011 site:wikipedia.org/wiki/	Watanabe, H.; Kanou, K. ; Kobayashi, Y. ; Fujie, M.G. Development of a "steerable drill" for ACL reconstruction to create the arbitrary trajectory of a bone tunnel. 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. Page(s): 955 – 960.
Dlib - Wikipedia	Rodriguez Mason Srinivasa "Abort and retry in grasping" 2011 site:wikipedia.org/wiki/	Rodriguez, Alberto, et al. "Abort and retry in grasping." Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on. IEEE, 2011. Rodriguez, A.; Mason, M. T.; Srinivasa, S. S.; Bernstein, M.; Zirbel, A. (2011). "Abort and retry in grasping". 2011 IEEE
Voronoi diagram - Wikipedia	Van Toll Cook Geraerts "Navigation meshes for realistic multi-layered environments" 2011 site:wikipedia.org/wiki/	van Toll, Wouter G.; Cook IV, Atlas F.; Geraerts, Roland (2011), Navigation Meshes for Realistic Multi-Layered Environments (PDF), International Conference on Intelligent Robots and Systems, IEEE/RSJ, pp. 3526–3532.
Routing Protocol for Low power and Lossy Networks – Wikipedia	Clausen Herberg Philipp "A critical evaluation of the IPv6 Routing Protocol" 2011 site:wikipedia.org/wiki/	Clausen, T.; Herberg, U.; Philipp, M.; "A critical evaluation of the IPv6 Routing Protocol for Low Power and Lossy Networks (RPL)", Wireless and Mobile Computing, Networking and Communications (WiMob), 2011 IEEE 7th International Conference on , vol., no., pp.365-372, 10-12 Oct. 2011
Distributed file system for cloud - Wikipedia	Kobayashi Mikami Kimura "The gfarm file system on compute clouds" 2011 site:wikipedia.org/wiki/	Kobayashi, K; Mikami, S; Kimura, H; Tatebe, O (2011). The Gfarm File System on Compute Clouds. Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011
Rétro-ingénierie en informatique — Wikipédia	Antunes Neves Verissimo "Reverse engineering of protocols from network traces" 2011 site:wikipedia.org/wiki/	(en) J. Antunes, N. Neves et P. Verissimo, « Reverse Engineering of Protocols from Network Traces », Reverse Engineering (WCRE), 2011 18th Working Conference on, 2011
Renren - Wikipedia	Yang Wilson Wang "Uncovering social network sybils in the wild" 2011 site:wikipedia.org/wiki/	Zhi Yang; Christo Wilson; Xiao Wang; Tingting Gao; Ben Y. Zhao; Yafei Dai (November 2011). "Uncovering Social Network Sybils in the Wild". Proc. of Internet Measurement Conference (IMC). Berlin, Germany.

Table 5.14. Manual data checks of Wikipedia citations for Industrial & Manufacturing Engineering (conference papers).

Wiki page	Query	Citation in Wikipedia
SRM Engine Suite - Wikipedia	Coble Smallbone Bhave "Implementing detailed chemistry and in-cylinder stratification into 0" 2011 site:wikipedia.org/wiki/	Coble; et al. (2011). "Implementing Detailed Chemistry and In-Cylinder Stratification into 0/1-D IC Engine Cycle Simulation Tools". SAE Technical Paper. doi:10.4271/2011-01-0849. SAE 2011-01-0849.
太陽光発電 - Wikipedia	Woodhouse James Margolis "An economic analysis of photovoltaics versus traditional energy" 2011 site:wikipedia.org/wiki/	An Economic Analysis of Photovoltaics Versus Traditional Energy Sources: Where Are We Now and Where Might We Be in the Near Future? Michael Woodhouse, Ted James, Robert Margolis, David Feldman, Tony Merkel, Alan Goodrich, NREL/CP-6A20-50714, July 2011
SORCER - Wikipedia	Li Feng Liu "A SOOA based distributed computing mechanism for road" 2011 site:wikipedia.org/wiki/	Li, Nan; Tao Feng; Bin Liu (2011). "A SOOA Based Distributed Computing Mechanism for Road Traffic Noise Mapping". IEEE Computer Society Washington, DC, USA: 109–112
SRM Engine Suite - Wikipedia	Smallbone Bhave Coble "Identifying optimal operating points in terms of engineering" 2011 site:wikipedia.org/wiki/	Smallbone; et al. (2011). "Identifying Optimal Operating Points in Terms of Engineering Constraints and Regulated Emissions in Modern Diesel Engines"
SRM Engine Suite - Wikipedia	Smallbone Bhave Coble "Simulating PM emissions and combustion stability in gasoline" 2011 site:wikipedia.org/wiki/	Smallbone; et al. (2011). "Simulating PM Emissions and Combustion Stability in Gasoline/Diesel Fuelled Engines". SAE Technical Paper. doi:10.4271/2011-01-1184. SAE 2011-01-1184.
Selektiv katalytisk reduktion – Wikipedia	Saedlou Santacreu Leseux "Suitable stainless steel selection for exhaust line containing" 2011 site:wikipedia.org/wiki/	Saedlou, Santacreu and Leseux (2011). Suitable Stainless Steel Selection for Exhaust Line Containing a Selective Catalytic Reduction (SCR). SAE International. sid. 1.
Active suspension - Wikipedia	Bryant Beno Weeks "Benefits of electronically controlled active electromechanical suspension systems" 2011 site:wikipedia.org/wiki/	Bryant, A.; Beno, J.; Weeks, D. (2011). "Benefits of Electronically Controlled Active Electromechanical Suspension Systems (EMS) for Mast Mounted Sensor Packages on Large Off-Road Vehicles"

Table 5.15. Manual data checks of Wikipedia citations for Building & Construction Engineering (conference papers).

Wiki page	Query	Citation in Wikipedia
Панагия ту Синти — Википедия	Philokyprou "The impact of different philosophical approaches towards the" 2011 site:wikipedia.org/wiki/	Philokyprou M., Petropoulou E. The ... The impact of different philosophical approaches towards the conservation of ancient monasteries in Cyprus // Structural ...
Transport Integration Act 2010 - Wikipedia	Pearce "The Transport Integration Act 2010 Driving integrated and" 2011 site:wikipedia.org/wiki/	Robert Pearce and Ian Shepherd ... Act 2011 amended the Transport Integration Act to reverse the merger of the Port of Melbourne Corporation and the Port of ...
Transport Legislation Review - Wikipedia	Pearce "The Transport Integration Act 2010 Driving integrated and" 2011 site:wikipedia.org/wiki/	Pearce and Shepherd observed that it ' ... The Government established a Taxi Industry Inquiry in May 2011 under a new agency, the Taxi Services Commission.
Neoweb - Wikipedia	Thakur Han Leshchinsky "Creep deformation of unreinforced and geocell-reinforced recycled asphalt" 2011 site:wikipedia.org/wiki/	Neoweb is a sustainable solution for road construction as it enables the use of locally available but marginal soils for infill, ... et al. 2011 and 2009). ...
Talk:Cellular confinement - Wikipedia	Thakur Han Leshchinsky "Creep deformation of unreinforced and geocell-reinforced recycled asphalt" 2011 site:wikipedia.org/wiki/	Talk:Cellular confinement WikiProject Soil ... (Han, et al. 2011). ... (Thakur, et al, 2010) ...
Neoweb - Wikipedia	Leshchinsky "Enhancing ballast performance using geocell confinement" 2011 site:wikipedia.org/wiki/	Neoweb is a sustainable solution for road construction as it enables the use of locally available but marginal soils for infill, ... et al. 2011 and 2009). ...

5.6 Wikipedia citations and Scopus citation counts (RQ8, RQ9)

For conference papers, 0.42% of the papers in Computer Science Applications have at least one Wikipedia citation, 1.35% of the papers in Software Engineering have at least one Wikipedia citation, 0.12% of the papers in Industrial & Manufacturing Engineering have at least one Wikipedia citation and 0.23% of the paper in Building & Construction Engineering have at least one citation from Wikipedia (Table 5.16). These proportions are too low to allow Wikipedia citations to be used to assess the impact of individual articles for most purposes, but they could still be used to compare the impact of groups of articles using proportion cited indicators (Thelwall, 2017a). For such a calculation to be credible, however, additional evidence would be needed of the value of Wikipedia citations in engineering due to the absence of correlation evidence in three of the four fields analysed here.

The low proportions of journal articles and conference papers were cited by Wikipedia (Table 5.16) could be due to low coverage of the subject areas chosen in Wikipedia, although it seems more likely that most articles and papers in these fields are not useful for Wikipedia.

Table 5.16. Average numbers of Wikipedia and Scopus citations (geometric mean) and percentage cited for both conference papers and journal articles.

Scopus subject category	**Journal articles**		**Conference papers**	
	Wikipedia citations Geometric mean (% cited)	**Scopus citations Geometric mean (% cited)**	**Wikipedia citations Geometric mean (% cited)**	**Scopus citations Geometric mean (% cited)**
Computer Science Applications	0.19459 (6.68%)	3.46 (80.9%)	0.00326 (0.42%)	0.53 (34.4%)
Software Engineering	0.01197 (1.36%)	3.43 (80.6%)	0.00119 (1.35%)	1.29 (54.7%)
Industrial & Manufacturing Engineering	0.00429 (0.52%)	2.54 (71.3%)	0.00093 (0.12%)	0.18 (17.5%)
Building & Construction Engineering	0.00706 (0.83%)	2.43 (71.7%)	0.00205 (0.23%)	0.20 (18.3%)

There are statistically significant correlations between Wikipedia citation counts and Scopus citation counts only in Computer Science Applications (Table 5.17). In the other three subject categories, the correlations are not statistically significantly different from zero. In Computer Science Applications, the correlation is higher for conference papers than for journal articles but in both cases the correlations are low.

Table 5.17. Spearman correlations between Wikipedia citations, and Scopus citations for articles, and conference papers in Scopus from 2011 in four engineering subjects.

Scopus subject category	Articles	Conf. papers	Spearman correlation for conference papers (95% CI)	Spearman correlation for journal articles (95% CI)
Computer Science Applications	5912	6700	0.274** (0.2517, 0.2960)	0.056** (0.0306, 0.0814)
Software Engineering	8231	8085	-0.001 (-0.0228, 0.0208)	0.020 (-0.0016, 0.0416)
Building & Construction Engineering	8406	1753	0.040 (-0.0068, 0.0867)	0.002 (-0.0193, 0.0234)
Industrial & Manufacturing Engineering	7354	5650	-0.017 (-0.4306, 0.0091)	0.012 (-0.0109, 0.0348)

**Statistically significant at p=0.01

5.7 Google Books citations data checks (RQ10)

Manual data checks of 25 random samples for the four subject categories (Table 5.18; see the following tables for details) show that only Building & Construction Engineering has all the correct citation text matches to all the queries, but all are accurate enough to be useful in practice.

Table 5.18. The results of manual checks of the Google Books search citations to conference papers of the 25 random citations in each subject category.

Scopus subject category	Citations	Correct Citations
Computer Science Applications	25	24(96%)
Software Engineering	25	22(88%)
Industrial & Manufacturing Engineering	25	24 (96%)
Building & Construction Engineering	25	25 (100%)

Table 5.19. Manual data checks for Google Books queries and the corresponding citation text, if found, for Industrial & Manufacturing Engineering.

Query	**Citation in Google Books page, if any**
Ali Mashor Mohd "A portable continuous blood pressure monitoring kit" 2011	Ali Hassan, M.K., Mashor, M.Y., Mohd Saad, A.R., Mohamed, M.S.: A Portable Continuous Blood Pressure Monitoring Kit. In: 2011 IEEE Symposium on Business
Guo Liu Song "A simple fast jacket transform for DFT based on generalized" 2011	Guo, Y., Liu, Y., Song, X.: A simple fast jacket transform for DFT based on generalized prime factor decomposing algorithm. In: IEEE Symposium on Business, Engineering and Industrial Applications, pp. 265–270 (2011)
Chang Chen Tsai "A new measurement method for power signatures of non-intrusive demand" 2011	Chang, H.H., Chen, K.L., Tsai, Y.P., Lee, W.J.: A New Measurement Method for Power Signatures of Non Intrusive Demand... 2196–2201 (2011)
Sun Biller Gu "Energy consumption reduction for sustainable manufacturing systems considering machines with" 2011	Sun, Z., Biller, S., Gu, F., Li, L.: Energy Consumption Reduction for Sustainable Manufacturing Systems Considering Machines with Multiple Power States.
Deshpande "Legacy machine monitoring using power signal analysis" 2011	A. Deshpande, R. Pieper, Legacy machine monitoring using power signal analysis, in: Proc. ASME 2011 Int. Manuf. Sci. Eng. Conf., 2011, pp. 1–8.
Hicks "Integrating a reheat steam cycle power or recovery boiler into" 2011	Hicks, T.E., 2011. Integrating a reheat steam cycle power or recovery boiler into an existing pulp mill. Portland, Oregon. Paper presented at the TAPPI Pulping, Engineering, Environmental, Recycling and Sustainability Conference
Rosa Rovida Viganò "Proposal about the use of data base in engineering design" 2011	**No corresponding citation**

Table 5.20. Manual data checks for Google Books queries and the corresponding citation text, if found, for Computer Science Applications.

Query	Citation in Google Books page, if any
Kim "Combined visually and geometrically informative link hypothesis for pose-graph visual" 2011	Kim, A., Eustice, R.M.: Combined visually and geometrically informative link hypothesis for pose-graph visual slam using bag-of-words. In: Proceedings of the IEEE/RSJIROS, San
Hunt Bachmann Murphy "A rapidly reconfigurable robot for assistance in urban search and" 2011	Hunt, A. J., R. J. Bachmann, R. R. Murphy and R. D. Quinn. 2011. A rapidly reconfigurable robot for assistance in urban search and rescue.
Das Maughan McCann "Towards mixed-initiative multi-robot field experiments Design deployment and lessons learned" 2011	Das J, Maughan T, McCann M, Godin M, O'Reilly T, Messie M, Bahr F, Gomes K, Py F, Bellingham J, ... Rajan K (2011) Towards mixed-initiative, multi-robot field experiments: design, deployment, and lessons learned.
Melendez-Calderon Bagutti Pedrono "Hi5 A versatile dual-wrist device to study human-human interaction and" 2011	Melendez-Calderon, A., Bagutti, L.,Pedrono, B., Burdet, E.: Hi5: a versatile dual-wrist device to study human-human interaction and bimanual control. In: 2011 IEEE/RSJ International Conference ...
Elbrechter Haschke Ritter "Bi-manual robotic paper manipulation based on real-time marker tracking and" 2011	Elbrechter C., Haschke R., and Ritter, H. Bi-manual robotic paper manipulation based on real-time marker tracking and physical modelling. In IEEE/RSJ International Conference on Robots and Systems (IROS), pp. 1427–1432, 2011.
Jonsson "Road condition discrimination using weather data and camera images" 2011	Jonsson, P. 2011a. ... In 2011 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA), September 19–21, ... Road condition discrimination using weather data and camera images.
Daly Ma Waslander "Coordinated landing of a quadrotor on a skid-steered ground vehicle" 2011	**No corresponding citation**

Table 5.21. Manual data checks for Google Books queries and the corresponding citation text, if found, for Computer Software.

Query	**Citation in Google Books page, if any**
Lee Fraundorfer Pollefeys "RS-SLAM RANSAC sampling for visual FastSLAM" 2011	Lee, G.H., Fraundorfer, F., Pollefeys, M.: RS-SLAM: RANSAC sampling for visual FastSLAM. In: International Conference on Intelligent Robots and Systems, pp. 1655–1660 (2011) Image-Based Smoke Detection in
Kaneko Kanehiro Morisawa "Humanoid robot HRP-4 Humanoid robotics platform with lightweight and slim" 2011	Kaneko, K., Kanehiro, F., Morisawa, M., Akachi, K., Miyamori, G., Hayashi, A., Kanehira, N.:Humanoid robot HRP-4 - humanoid robotics platform with lightweight and slim body. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.
Padoy "3D thread tracking for robotic assistance in tele-surgery" 2011	**No corresponding citation**
Daly Ma Waslander "Coordinated landing of a quadrotor on a skid-steered ground vehicle" 2011	**No corresponding citation**
Zhang "Hybrid ant colony optimization based on genetic algorithm for container" 2011	Zhang DZ, Du LN (2011) Hybrid ant colony optimization based on genetic algorithm for container loading problem. In: Proceedings of the international conference of soft computing and pattern recognition, pp. 890–900 Abbattista
Waldmann "There's never enough time Doing requirements under resource constraints and" 2011	Waldmann, B.: There's never enough time: doing requirements under resource constraints, and what ...
Klenner "An incremental entity-mention model for coreference resolution with restrictive antecedent" 2011	Manfred Klenner. Enforcing ... An incremental entity-mention model for coreference resolution with restrictive antecedent accessibility. In Proc. of the ... Hissar, Bulgaria, 2011. Cited on ...

Table 5.22. Manual data checks for Google Books queries and the corresponding citation text, if found, for Building & Construction Engineering.

Query	Citation in Google Books page, if any
Klein Kavulya Jazizadeh "Towards optimization of building energy and occupant comfort using multi-agent" 2011	Klein, L., Kavulya, G., Jazizadeh, F., Kwak, J., Becerik-Gerber, B., Varakantham, P., Tambe, M.:Towards optimization of building energy and occupant comfort using multi-agentsimulation. In: The 28th International Symposium on Automation ...
Beller "Development of a simulation tool to predict urban wind potential" 2011	Beller, C. (2011). Development of a simulation tool to predict Urban wind potential. In Sustainability in energy and buildings, smart innovation, systems and technologies (Vol. 7, pp. 111–120). Springer, Berlin. Balduzzi, F., Bianchini, A., ...
Fuchida "Evaluation of response characteristics of buried pipelines during earthquakes" 2011	Fuchida, K., Evaluation of response characteristics of buried pipelines during earthquakes, Proc. of 8th Int. Conf. on Earthq. Resist. Eng. Struct., pp. 163–172, 2011.
Gomes "Metro line implementation in a European city" 2011	N. M. Gomes. Rocha. PROEC – Projectos, Estudos e Construções Lda., Braga, ... www.witpress.com, ISSN 1743-3509 (on-line) Urban Transport XVII 393 Metro line implementation in a European city.
Claesson "An analytical method to calculate borehole fluid temperatures for time-scales" 2011	Claesson, J., Javed, S., 2011. An analytical method to calculate borehole fluid temperatures for time-scales from minutes to decades. ASHRAE Transactions 117, 279e288. Cui, P., Li, X., Man, Y., Fang, Z.H. ...
Cysewska-Sobusiak Hulewicz Krawiecki "Examples of the application of light-tissue interaction to biomedical engineering" 2011	A. Cysewska-Sobusiak, A. Hulewicz, Z. Krawiecki & G. Wiczynski Poznan University of Technology, Institute of Electrical ... Architecture and the Environment 223 Examples of the application of light-tissue interaction to biomedical engineering.
Ching Chen Phoon "Updating uncertainties in friction angles of clean sands" 2011	Ching, J., Chen, J.R., Yeh, J.Y. & Phoon, K.K. (2012) Updating uncertainties in friction angles of clean sands. ASCE Journal of Geotechnical and Geoenvironmental Engineering, 138 (2), 217–229. Christian, J.T. & Baecher, G.B. (2011) ...

5.8 Google Books citations and Scopus citation counts for conference papers (RQ11, RQ12)

For conference papers, 15.9% of the papers in Computer Science Applications have at least one Google Books citation, 35.8% of the papers in Software Engineering have at least one Google Books citation, 1.8% of the papers in Industrial & Manufacturing Engineering have at least one Google Books citation and 17.2% of the papers in Building & Construction

Engineering have at least one Google Books citation (Table 5.23). Since in all areas most papers have few Google Books citations, this data source is inadequate to assess the impact of typical books but may still be used to find highly cited papers or to compare the impacts of groups of papers.

Table 5.23. Average numbers of Google Books and Scopus citations (geometric mean) and percentage cited for conference papers.

Scopus subject category	Geometric mean (% cited)	
	Google Books citations	Scopus citations
Computer Science Applications	(0.158) 15.9%	(0.690) 37.2%
Software Engineering	(0.473) 35.8%	(1.602) 58.5%
Industrial & Manufacturing Engineering.	(0.015) 1.8%	(0.273) 23.1%
Building & Construction Engineering.	(0.148) 17.2%	(0.313) 25.2%

There are statistically significant correlations between Google Books citation counts and Scopus citation counts in Computer Science Applications, Software Engineering and Industrial & Manufacturing Engineering but not in Building & Construction Engineering (Table 5.24).

Table 5.24. Spearman correlations between Google Books citations, and Scopus citations (95% confidence interval) for conference papers in Scopus from 2011 in four engineering subjects.

Scopus subject category	Papers	Spearman correlation (95% CI)
Computer Science Applications	9982	0.305** (0.2871, 0.3227)
Software Engineering	9990	0.285** (0.2669, 0.3029)
Industrial & Manufacturing Engineering	9990	0.092** (0.0725, 0.1114)
Building & Construction Engineering	4602	0.000 (-0.0289, 0.289)

**Statistically significant at p=0.01

5.9 Summary

This chapter gave results from prevalence statistics, Spearman correlations (with Scopus citations) and outlier analyses for Mendeley readers for both journal articles and conference papers in four subject categories. It also gave results from accuracy checks, prevalence statistics, and Spearman correlations (with Scopus citations) for three web indicators: Google Patents citations, Wikipedia citations and Google Books citations for conference papers and journal articles on the same subject categories.

All the methods used were accurate for conference papers in all fields, with the partial exception of Google Books, which had 12% incorrect citations in one subject area.

Substantial minorities or small majorities or papers were read in Mendeley and substantial minorities had Google Books citations in all fields but tiny percentages had Wikipedia or Google Patents citations (Table 5.25).

Table 5.25. Percentages cited in (Conference papers) for Mendeley readers, Google Patents citations, Wikipedia citations and Google Books citations for four engineering fields.

Scopus subject category	Mendeley readers	Google Patents Citations	Wikipedia Citations	Google Books citations
Computer Science Applications	47.2%	0.31%	0.42%	15.9%
Computer Software Engineering	68.6%	0.22%	1.35%	35.8%
Building & Construction Eng.	18.7%	0%	0.23%	17.2%
Industrial & Manufacturing Eng.	41.0%	0.11%	(0.12%	1.8%

**Statistically significant at p=0.01

There were moderate correlations between Mendeley readers and Scopus citation counts for both journal articles and conference papers in all the subject areas (Table 5.26). There were significant correlations in the three subject areas in Google Books citations except in Building & Construction Engineering with no correlations. Whereas, in Wikipedia and Google Patents citations, the correlations were significant only in one subject category (Computer Science Applications).

Table 5.26. Spearman correlations (conference papers) for Mendeley readers, Google Patents citations, Wikipedia citations and Google Books citations for four engineering fields.

Scopus subject category	Mendeley readers	Google Patents Citations	Wikipedia Citations	Google Books citations
Computer Science Applications	0.439**	0.072**	0.274**	0.305**
Computer Software	0.437**	0.018	-0.001	0.285**
Building & Construction Eng.	0.143**	0.000	0.040	0.000
Industrial & Manufacturing Eng.	0.168**	-0.015	-0.017	0.092**

**Statistically significant at p=0.01

Chapter 6. Discussion

This chapter compares the findings to the most relevant prior research and discuses reasons for the results as well as their wider implications.

6.1 Introduction

This thesis investigates indicators for the impact of conference papers and journal articles using four different types of web indicators: Mendeley readers, Google Patents citations, Wikipedia citations and Google Books citations. The thesis analyses journal articles and conference papers in four selected engineering fields, comparing the web indicators with Scopus citation counts. The thesis also assesses the accuracy of methods to count citations from Google Patents, Wikipedia and Google Books to conference papers. It also investigates why articles in Mendeley are widely read but have few Scopus-indexed citations, or are highly cited conference papers with few Mendeley readers.

6.2 Data collection accuracy (RQ4, RQ7, RQ10)

RQ4: Can citations from Google Patents to conference papers be automatically extracted using curated Bing queries?

RQ7: Can Wikipedia citations to conference papers be automatically extracted using curated Bing queries?

The methods to collect Google Patents citations and Wikipedia citations through Bing queries did not return false matches. Thus, incorrect matches are not a major concern for matching conference papers with Bing queries. This is a reassuring finding since it was possible that references to conference papers might be less precise than references to journal articles, for example if scholars published similar papers in conferences or alternative versions in conferences and journal articles. It is also reassuring because web search engines evolve over time and there has been a tendency in recent years for search engine web interfaces to return results that do not technically match the user query but are relevant in another way. This did not occur with any of the Bing API queries, perhaps because they were complex or the extra results are not provided through the API interface.

RQ10: Can citations from Google Books to conference papers be automatically extracted using Google Books API queries and heuristics to filter the results?

The methods used to collect Google Books citations through Google Books API queries and subsequent filtering found a small proportion of false matches, but less than 10% overall. Thus, Google Books citations should be used with some care in case of false matches, but the accuracy rate seems to be high enough to be used in practice. Prior studies have found matching issues due to books with multiple volumes or editions, but this does not seem to be a problem for conferences.

6.3 Mendeley readers for conference papers (RQ1, RQ2)

RQ1: Do Mendeley readership counts reflect the scholarly impact of conference papers in conference-based fields?

Mendeley readership counts and Scopus citation counts have strong and significant positive correlations for journal articles in all the engineering fields analysed and for conference papers in the two computing fields but not in the other two engineering fields, Industrial and Manufacturing Engineering and Building and Construction Engineering, which have weak but positive correlations.

For conference papers, 68.6% of the papers in the Computer Software subject category have at least one Mendeley reader and 54.7% of the papers have at least one Scopus citation. These findings show that the impact of conferences is high in Scopus and Mendeley for computing research. In Building & Construction Engineering, conference papers have much lower percentage coverage; 18.3% and 18.7% of the papers have at least one Scopus citation and at least one Mendeley reader, respectively. This may be due to low coverage of conference proceedings in the field of engineering for Scopus but, this cannot explain the results for Mendeley. It may be that a high percentage of engineering conference papers are not of interest to publishing academics, either because of their applied focus or due to disciplinary norms in citation practices.

If Mendeley readers and citations are not useful in some fields, then new indicators, perhaps including download counts, would be needed to reflect this impact. Unfortunately, counting article downloads gives an imperfect measure of readership because someone might decide not to read an article after downloading it and accessing its abstract; while others might read an article without accessing it electronically because they subscribe to a print version of the journal, whereas they can read it in a library or were given a printout of the article by their lecturer. Also, there are technical problems with downloads counts. Some article downloads may be from web crawlers or other computerised processes that do not reflect human readers and more so, people may download an article multiple times despite reading it only once. Although this may occur by accident or because of not saving a local copy and needing to check a document several times.

RQ2: Does the answer to the above research question differ between fields in comparison to journal articles?

For journal articles, the strong and positive correlations between Mendeley readership and citation counts for all four of the studied Engineering subject categories corroborate past studies of other areas (Li, Thelwall, & Guistini, 2012; Bar-llan, 2012; Mohammadi & Thelwall, 2014). The findings broadly agree with the study of Zahedi et al. (2015) across 5 major fields in science which reported correlations between Mendeley and citations counts: Social sciences and Humanities (0.614), Natural Sciences and Engineering (0.597), Life and Earth sciences (0.578), Biomedical and Health sciences (0.553), and Mathematics and Computer sciences (0.457). Also, the study corroborates Mohammadi et al. (2015), which reported moderate correlations between Mendeley readers and citation counts in Clinical

Medicine (0.463), Engineering and Technology (0.327), social sciences (0.456), Physics (0.308) and Chemistry (0.369). This research findings also corroborates the study of Li et al. (2012) for Nature and Science articles. For Nature articles, the study found high correlations (0.559) between Mendeley readers and Web of science citation counts and high correlations (0.592) between Mendeley readers and Google scholar citation counts. For Science articles, Mendeley readers have moderate correlations (0.540) with Web of Science citation counts and Mendeley readers have high correlations (0.603) with Google Scholar citation counts.

In summary, journal article reader counts correlated more strongly with Scopus citations than did conference papers. The higher correlations for journal articles may be due to Scopus indexing fewer conference proceedings and therefore missing out a greater proportion of citations to conference papers. This assumes that conference papers are more likely to be cited by other conference papers than by journal articles.

6.4 Causes of outliers in Mendeley readers and citation counts for conference papers

RQ3: What are the causes of conference papers having many Mendeley readers compared to citations or many citations compared to Mendeley readers?

The causes of higher Mendeley readership counts than citation counts or vice versa are based on the small sample of papers analysed in the current study and may not apply to other years or fields. The list of reasons why papers may attract many readers compared to their citations, or the other way round, shows that there are several legitimate causes of outliers. It is therefore important to accept that Mendeley reader counts will not always be a good approximation to Scopus citation counts for individual papers.

The possible causes of outliers add to those found in Thelwall (2017b) for journal articles.

a) Articles that are useful for students, who do not publish papers in Scopus.
b) Articles that are useful for professionals that do not author articles.
c) Papers with general interest attract readers from community that do not use the article in their work.
d) Articles from countries whose researchers do not publish in Scopus journals.

Thelwall (2017b) also identified some reasons for articles having many Scopus citations compared to their Mendeley readers as follows:

a) Articles from an academic community that do not use Mendeley due to limited access to an internet.
b) Articles mainly from a publishing author community that cannot access Mendeley.
c) Updated articles: Users of Mendeley may register as readers of an update of an article rather than the original version thus, authors may cite the original.
d) Multidisciplinary articles that tend to attract many citations from one side of their focus with few readers per citation in a field norm legitimately classified with another category.

In addition, Thelwall (2017b) identified some reasons for outliers due to technical limitations.

a) Inclusion of irrelevant journal in a Scopus category from fields with different ratios of reader counts to citation counts
b) Articles with missing or incorrect DOIs in Scopus, Mendeley, or with multiple valid DOIs, with different versions in both Mendeley and Scopus.
c) Articles that appear more than once in the Scopus tend to reduce the Scopus-indexed citation counts of both versions

Given the wide range of valid reasons for outliers between citation counts and reader counts, the new ones found in this thesis should be seen as adding to the existing set and not being necessarily specific to the fields or document types examined. In particular, the following new reasons may be of general relevance within academia, in terms of papers that may have higher Mendeley than Scopus impact.

- Papers that create public awareness, motivation and participation for new scientific discoveries
- Papers that are relevant to daily life or religious beliefs.
- Practical solutions to important real-world problems.
- Social media articles of general interest to users.
- Practical commercial advice.

Similarly, new reasons for relatively high citation counts compared to readers include the following.

- Papers on software packages
- Papers with a set model for completing a task

6.5 Google Patents citations for conference papers (RQ5, RQ6)

RQ5: Are there enough citations from Google Patents to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Google Patents in conference-based fields?

The four selected engineering fields have very low Google Patents citation counts for both conference papers and journal articles. For conference papers, 0.3% of the papers in Computer Science Applications have at least one Google Patents citation and 34.4% of the papers have at least one Scopus citation. These findings show that the impact of conferences is very low for Google Patents citations and high for Scopus citations. In Building & Construction Engineering, conference papers have no Google Patents citations and 18.3% have at least one Scopus citation.

RQ6: Do citations to conference papers from Google Patents reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Google Patents in conference-based fields?

Google Patents citation and Scopus citation counts have positive significant correlations for journal articles and conference papers for Computer Science Applications. In Software Engineering, Industrial & Manufacturing Engineering and Construction Engineering, these

categories have correlations that are not statistically significantly different from zero. This might be due to low numbers of patent citations, which makes it difficult to obtain a high correlation (Thelwall, 2016b).

The positive correlation between Google Patents citations and Scopus citation counts for Computer Science Applications journal articles (0.210), broadly agreeing with the study of Kousha and Thelwall (2015c) where journal articles in Scopus subject categories for sixteen applied Science and Engineering fields were analysed: in Biomedical Engineering, 10.1% of the articles have at least one Google Patents citations; as well as Biochemistry & Molecular Biology (5.4%); Biotechnology (9.2%); Chemical engineering (2.9%); Computer Science (5.9%); Control & Systems Engineering (3.9%); Electrical & Electronic Engineering (5.6%); Energy Engineering (2.2%); Environmental Engineering (2.7%); Food science (5.5%); Industrial & Manufacturing Engineering (2.5%); Mechanical Engineering (1.9%); Pharmacology & Pharmaceuticals (6.8%); Physics Instruments & Instrumentations (2.8%); Polymer science (4.6%); and Surgery (2.9%). The same study found week to moderate (0.005 to 0.36) positive correlations with Scopus citation counts across all the applied science and engineering fields analysed, which also corroborates with findings in this thesis.

This positive significant correlation between Google Patents citations and Scopus citations for Computer Science articles shows that patents citation might partly reflect scholarly impact in Computer Science.

6.6 Wikipedia citations for conference papers (RQ8, RQ9)

RQ8: Are there enough citations from Wikipedia to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Wikipedia in conference-based fields?

The four selected engineering fields have very low Wikipedia citation counts for both conference papers and journal articles. This study therefore shows that Wikipedia has very few citations to engineering fields.

This study found much lower values than Kousha et al. (2016) for books, where in Arts 58% to 61% of books have at least one Wikipedia citation, as well as Humanities (48% to 54%); Social Sciences (30 to 39%); Science (23 to 35%); and Engineering (18 to 37%). This shows that Wikipedia citations to conference papers are much less prevalent than citations to books.

For conference papers, 0.42% of the papers in Computer Science Applications have at least one Wikipedia citation, 1.35% of the papers in Software Engineering have at least one Wikipedia citation, 0.12% of the papers in Industrial & Manufacturing Engineering have at least one Wikipedia citation and 0.23% of the paper in Building & Construction Engineering have at least one citation from Wikipedia. These proportions are too low to allow Wikipedia citations to be used to assess the impact of individual articles for most purposes, but they could still be used to compare the impact of groups of articles using proportion cited indicators (Thelwall, 2017a). For such a calculation to be credible, however, additional

evidence would be needed of the value of Wikipedia citations in engineering due to the absence of correlation evidence in three of the four fields analysed here.

RQ9: Do citations to conference papers from Wikipedia reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Wikipedia in conference-based fields?

Wikipedia citation counts, and Scopus citation counts have statistically significant positive significant correlations for journal articles and conference papers for Computer Science Applications, but not in the other fields. Thus, overall, there is little evidence that Wikipedia citations reflect scholarly impact. This could be due to low numbers of Wikipedia citations, which makes it difficult to obtain a high correlation.

Overall, there is little evidence that Wikipedia citations reflect scholarly impact. This could be due to low numbers of Wikipedia citations, which makes it difficult to obtain a high correlation. Also, there were low proportions for Wikipedia citations to be used to assess the impact of individual articles for most purposes, but they could still be used to compare the impact of groups of articles using proportion cited indicators (Thelwall, 2017a).

In Software Engineering, Industrial & Manufacturing Engineering and Building & Construction Engineering, these categories have correlations that are not statistically significantly different from zero. This might be due to low numbers of Wikipedia citations, which makes it difficult to obtain a high correlation (Thelwall, 2016b).

6.7 Google Books citing conference papers

RQ11: Are there enough citations from Google Books to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Google Books in conference-based fields?

For conference papers, 15.9% of the papers in Computer Science Applications have at least one Google Books citation, 35.8% of the papers in Software Engineering have at least one Google Books citation, 1.8% of the papers in Industrial & Manufacturing Engineering have at least one Google Books citation and 17.2% of the papers in Building & Construction Engineering have at least one Google Books citation. Since in all areas most papers have few Google Books citations, this data source is inadequate to assess the impact of books citations but may still be used to find highly cited papers or to compare the impacts of groups of papers.

RQ12: Do citations to conference papers from Google Books reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Google Books in conference-based fields?

Google Books citations and Scopus citation counts have statistically significant positive correlations for conference papers for Computer Science Applications (0.305), Software Engineering (0.285) and Industrial & Manufacturing Engineering (0.092) but not in Building & Construction Engineering. Google Books citations have previously been shown to reflect

scholarly impact for journal articles (Kousha, Thelwall & Rezaie, 2011), where Google Books citations have positive Spearman correlations with WoS citations in Sciences (0.15 to 0.35), Social sciences (0.14 to 0.59) and Humanities (0.35 to 0.65), and the current study extends this to conference papers in some conference-based engineering fields.

In Building & Construction Engineering, there is no correlation between Google Books citation and Scopus citation counts; this might be due to low coverage of Google Books in this field.

6.8 Limitations

The results of this thesis are limited by several factors. The results were only analysed for four fields and there may be different results for other fields that value conferences. The results also cover only one year and the value of conferences may change over time, so could be different in 2018 and beyond. The results are also restricted to the subject delimitation of Scopus and so some important conferences may have been excluded. The use of Scopus also precluded the analysis of small fields, such as computational linguistics, that do not have their own Scopus category. Different results may also have been obtained if using citation counts from the Web of Science or another database rather than Scopus. Probably the scholarly databases with the widest coverage of conference citations now include Google Scholar, Microsoft Academic and Dimensions but only the first of these was available when the data was collected and this did not allow automated queries.

A more generic limitation is that this thesis only investigates citations to conference papers in engineering fields in which they seem to be important. Thus, the two positive results (Mendeley readers and Google Books citations) should not be extrapolated to fields in which conferences are less important. Conversely, it seems reasonable to assume that Google Patents citations and Wikipedia citations would have little value for conference papers in nearly all fields.

The correlation tests used are limited as methods to identify whether an indicator reflects scholarly impact because they are an indirect method and do not address the reason why the citations or readers were created. They are only a first step towards validating an alternative indicator.

6.9 Summary

This chapter used Spearman correlations to analyse the strength of association between Mendeley readers and Scopus citation counts for conference papers and journal articles for four Engineering related fields. The study found moderate correlations between Mendeley readership counts and Scopus citation counts for both journal articles and conference papers. This chapter also compared citations from three web indicators with Scopus citation counts for conference papers for the same four fields. The findings suggest that Google Patents citations and Wikipedia citations have little value for conference papers in the selected Engineering related fields, but both Google Books citations and Mendeley readers are numerous enough to have some uses for conference papers. The results also include new

reasons for outliers in terms of the relationship between Mendeley reader counts and Scopus citation counts.

Chapter 7: Conclusions

7.1 Introduction

This study examined indicators for the impact of conference papers and journal articles using four different types of web indicators: Mendeley readers, Google Patents citations, Wikipedia citations and Google Books citations. The study analysed journal articles and conference papers in four selected engineering fields (Computer Science Applications, Software Engineering, Building & Construction Engineering and Industrial & Manufacturing Engineering), comparing the web indicators with Scopus citation counts. This study also investigated why articles in Mendeley are widely read but have few Scopus-indexed citations, or highly cited but with few Mendeley readers, in conference papers. It also assessed the accuracy of methods to count citations from Google Patents, Wikipedia and Google Books to conference papers.

This chapter draws conclusions about the main findings for each of the research questions. Theoretical and empirical contributions of the research and research contributions to knowledge are then summarised. Finally, some recommendations for future research are given.

7.2 Answers to the research questions

RQ1: Do Mendeley readership counts reflect the scholarly impact of conference papers in conference-based fields?

Spearman correlations between Mendeley readership counts and Scopus citation counts for all the engineering related subject categories gave statistically significant moderate correlations between Mendeley readership counts and Scopus citations counts for conference papers in all the subject categories. Since citation counts are accepted as scholarly impact indicators, the correlations give some evidence that Mendeley readers reflect the scholarly impact of conference papers in all the subject categories. In conjunction with similar prior results for journal articles in all academic fields and triangulation with surveys and outlier analyses, it would be reasonable to use Mendeley reader counts as an academic impact indicator for conference papers in engineering fields.

RQ2: Does the answer to the above research question differ between fields in comparison to journal articles?

There were moderate correlations in comparing conference papers to journal articles in all the engineering fields. There were not substantial field or publication type (journal article, conference paper) differences in the apparent relationship between Mendeley readership counts and Scopus citations counts. Thus, the results do not point to disciplinary differences for the fields covered here, although Mendeley reader counts are likely be much less valuable for conference papers in non-engineering fields for which they are less important.

RQ3: What are the causes of conference papers having many Mendeley readers compared to citations or many citations compared to Mendeley readers?

The following attributes suggest that a paper may attract relatively many Mendeley readers for its Scopus citations: improving the performance of an existing system; creating public awareness, motivation and participation for new scientific discoveries; relevance to religious beliefs; practical solutions to important real-world problems; general interest; regional interest. These many different reasons relate to attracting a wider non-citing audience and add to the list of reasons found in previous research (including technical reasons). The list of reasons confirms that Mendeley sometimes has the ability to reflect non-academic impacts, albeit for a small minority of articles.

In contrast, papers that are highly cited compared to read may be about software packages or a set model for completing a task. These might be relatively perfunctory citations or citations by people that had not necessarily read the cited paper. Nevertheless, both reasons suggest that the cited paper represented useful methodological innovation and so does not point to citations over-valuing prior research.

RQ4: Can citations from Google Patents to conference papers be automatically extracted using curated Bing queries?

The results show that high precision queries are possible but the recall of these queries is unknown. The evidence is not strong, however, given the few results found. Much larger scale studies would be needed to give a more definitive answer to this question.

RQ5: Are there enough citations from Google Patents to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Google Patents in conference-based fields?

There were few citations from Google Patents to conference papers in Computer Science Applications, Computer software engineering and Industrial & Manufacturing Engineering. There were no citations from Google Patents to Building & Construction Engineering. These findings suggest that Google Patents is not useful for identifying the commercial impact of engineering research published in conferences. The results were not significantly better for journal articles in these fields. Although some indicator formulae, such as proportion cited indicators, can be used for altmetric data that mostly consists of zeros, the extreme scarcity of non-zero scores seems to rule out even this application.

RQ6: Do citations to conference papers from Google Patents reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Google Patents in conference-based fields?

The very weak correlations between Scopus citations and Google Patents citations to both journal articles and conference paper was too weak in almost all cases to give any evidence of the type of impact, if any, that Google Patents citations reflect. The partial exception is the weak correlation (0.210) for journal articles in Computer Science Applications. Whilst this gives a glimmer of hope that Google Patent citations may have value in some engineering areas, the overall answer to the research question is negative.

RQ7: Can Wikipedia citations to conference papers be automatically extracted using curated Bing queries?

High precision queries are possible but the recall of these queries is unknown. As for Google Patents citations, few citations were found and so larger scale evidence is needed to give confidence in this answer.

RQ8: Are there enough citations from Wikipedia to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Wikipedia in conference-based fields?

There were few citations from Wikipedia to conference papers or journal articles in all fields. Thus, Wikipedia seems to have little value for identifying the informational impact of individual articles or papers in engineering fields, even if using a proportion cited indicator.

RQ9: Do citations to conference papers from Wikipedia reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Wikipedia in conference-based fields?

The situation for Wikipedia exactly parallels that of Google Patents. The correlations between Scopus citations and Wikipedia citations to both journal articles and conference paper were too weak in almost all cases to give any evidence of the type of impact, if any, that Wikipedia citations reflect. The partial exception is the weak correlation (0.274) for journal articles in Computer Science Applications. Whilst it is possible that Wikipedia citations have value in some engineering areas, they are unlikely to have much value in most.

RQ10: Can citations from Google Books to conference papers be automatically extracted using Google Books API queries and heuristics to filter the results?

Reasonably high precision queries are possible but the recall of these queries is unknown. This is a more robust conclusion than for Google Patents and Wikipedia citations, given the much larger number of hits from the queries.

RQ11: Are there enough citations from Google Books to conference papers for altmetric purposes? Does this differ from the case for citations to journal articles from Google Books in conference-based fields?

There was significant coverage of Google Books citations to conference papers compared to Scopus citations in all the engineering subject categories. There is sufficient data from Google books to be able to separate out journal articles or conference papers by Google Books citation count scores.

RQ12: Do citations to conference papers from Google Books reflect a similar type of impact to Scopus citations? Does this differ from the case for citations to journal articles from Google Books in conference-based fields?

There were moderate correlations between Google Books citation counts and Scopus citation counts for conference papers in all the subject categories except Building & Construction

Engineering. This might be due to differences for Google Books citations even within the engineering subject categories. However, due to the significant correlations between Google Books citations and Scopus citations counts for conference papers in the three subject categories (Computer Science Applications, Computer Software Engineering and Industrial & Manufacturing Engineering), it is reasonable to use Google Books citations to conference papers as a scholarly impact indictor. Nevertheless, more evidence of this is needed to make this a robust conclusion, and to discover the source of the problem in one of the categories.

7.3 Research contributions

Several studies have found significant moderate correlations between Mendeley readership counts and citations counts for journal articles. No previous research has investigated the relationship between Mendeley readership counts and citation counts for conference papers in engineering fields. This thesis therefore gives the first empirical evidence that Mendeley reader counts have potential as an impact indicator for conference papers in engineering fields in which they are important.

Google Patents has been shown to be a useful source of patent citations to journal articles in some fields, but not for conference papers. This thesis provides evidence for the first time that Google Patents citation counts for conference papers have little value because they are too rare to be useful and may not reflect a useful type of impact, even in engineering fields that value conferences.

Previous studies have shown that Wikipedia citations can be used to help assess the knowledge transfer impact of the cited articles and books. No studies have investigated Wikipedia citations to conference papers. This thesis provides the first evidence that Wikipedia citations have little value for conference papers, even in engineering fields that value conferences.

There is substantial evidence from previous literature that Google Books citations reflect scholarly impact for journal articles and books but no evidence for conference papers. It may be particularly useful for conference-oriented engineering fields because journal articles are less central to these areas and so traditional journal-based citation indexes, such as Scopus and Web of Science, may not serve them well. This thesis provides the first evidence that Google Books citations to conference papers have value for engineering fields.

7.4 Research implications

This study has found evidence that Mendeley readership counts and Google Books citation counts for conference papers are reasonable impact indicators for research evaluation in the engineering fields for which conferences are important, although has not found any evidence that these indicators have more value for conference papers than for journal articles. In contrast, Wikipedia citations and Google Patents citations are not recommended as research indicators for conference papers in engineering fields. Thus, the most important practical implication of this thesis is that future research evaluations of researchers or teams that

include engineering research should consider two new indicators to support their work: Google Books citations for book-type impact and Mendeley readers for early impact.

7.5 Recommendations for future studies

Although conference papers are important in computer and engineering fields, due to the differences in the levels of activities of scholars across different disciplines, future studies are needed to explore other computing and engineering fields to assess whether the fields investigated in this thesis are unusual in any respect. Different publication years should also be investigated to check whether web citations are more popular for older or newer conference papers. For data triangulation purposes, it would also be helpful to use different citation databases, such as Dimensions, Scopus, Microsoft Academic and the Web of Science, to check whether they would give different results to those with the Scopus citation counts used in this thesis. It would also be useful to investigate whether other reference managers might give better results than Mendeley. Although this seems unlikely, it is possible that engineers use another reference manager that does not have wide uptake elsewhere in academia.

The relative lack of citations from Google Patents to conference papers is perhaps surprising given the applied nature of engineering subject areas. It is not clear whether the results are due to a tendency to avoid citing conference papers in references or whether the problem lies in the indexing of conference papers in the reference lists of patents by Google. Thus, future work may investigate the extent to which conference papers are cited in patents.

Future studies in web indicators for conference papers also need to go beyond the four selected indicators (Mendeley readership counts, Google Patents citations, citations from Wikipedia and Google Books citations) to attempt to capture other relevant aspects of the evaluated outputs. Syllabus mentions are an obvious choice but other social web indicators, such as from Twitter and Facebook, should also be assessed.

Efforts should also be made to learn how altmetrics should be conceptualized, calculated, aggregated, normalized, and standardized for conference papers. Many of these questions would benefit from the perspective of qualitative research into the motivations of different user groups to comment and share research products online. Quantitative research is also necessary to find the best and most reliable methods to calculate and normalize different altmetrics for research evaluation.

References

Abbas, A., Zhang, L. & Khan, S.U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37, 3-13.

Abrizah, A. & Thelwall, M. (2014). Can the impact of non-Western academic books be measured? An investigation of Google Books and Google Scholar for Malaysia. *Journal of the Association for Information Science and Technology*, 65(12), 2498-2508.

Adie, E. & Roe, W. (2013). Altmetric: enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, 26(1), 11-17.

Aduku, K.J., Thelwall, M., & Kousha, K. (2016). Do Mendeley reader counts reflect the scholarly impact of conference papers? An investigation of Computer Science and Engineering fields. In: Ismael Rafols, Jordi Molas-Gallart, Elena Castro-Martínez, Richard Woolley, *Proceedings of the 21th International conference on science and technology indicators – Peripheries, frontiers and beyond*, Valencia (Spain).

Aibar, E., Lladós-Masllorens, J., Meseguer-Artola, A., Minguillón, J., & Lerga, M. (2015). Wikipedia at university: what faculty think and do about it. *The Electronic Library*, 33(4), 668-683.

Alcácer, J. & Gittelman, M. (2006). Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations. *Review of Economics and Statistics*, 88(4), 774–779.

Ani, O. E. (2017). A bibliometrics analysis of the visibility of library and information science research in Nigeria in the Web of Science, 2000-2014. *African Journal of Library, Archives & Information Science*, *27*(1), 41–51.

Ardanuy, J. (2013). Sixty years of citation analysis studies in the humanities (1951–2010). *Journal of the Association for Information Science and Technology*, 64(8), 1751-1755.

Arrive, L., Boelle, P., Dono, P., Lewin, M., Monnier-Cholley, L. & Tubiana, J. (2004). Subsequent publication of orally presented original studies within 5 years after 1995 RSNA Scientific Assembly. *Radiology*, 232(1), 101-106.

Baker, D. (1978). Characteristics of the literature used by English musicologists. *Journal of Librarianship*, 10(3), 182–200.

Bar-Ilan, J. (2012). JASIST 2001–2010. *Bulletin of the Association for Information Science and Technology*, 38(6), 24-28.

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H. & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. *arXiv preprint arXiv:1205.5611*.

Becher, T. & Trowler, P. (2001). *Academic tribes and territories* (2ed). Milton Keynes, UK: Open University Press.

Boell, S.K., & Wilson, C.S. (2010). Journal Impact Factors for evaluating scientific performance: use of h-like indicators. *Scientometrics*, 82(3), 613–626.

Bollen, J., Van de Sompel, H., & Rodriguez, M.A. (2008). Towards usage-based impact metrics: First results from the mesur project. *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*. New York, NY: ACM Press (pp. 231-240).

Bollen, J., Rodriquez, M.A. & Van de Sompel, H. (2006). Journal status. *Scientometrics*, *69*(3), 669-687.

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PloS One*, 4(6), e6022.

Borrego, Á. & Fry, J. (2012). Measuring researchers' use of scholarly information through social bookmarking data: A case study of BibSonomy. *Journal of Information Science*, *38*(3), 297-308.

Bornmann, L. (2012). Measuring the societal impact of research: research is less and less assessed on scientific impact alone--we should aim to quantify the increasingly important contributions of science to society. *EMBO reports*, 13(8), 673-676.

Bornmann, L. (2014). Is there currently a scientific revolution in scientometrics? *Journal of the Association for Information Science and Technology*, 65, 647–648.

Bould, M. D., Hladkowicz, E. S., Pigford, A. E., Ufholz, L., Postonogova, T., Shin, E., & Boet, S. (2014). References that anyone can edit: Review of Wikipedia citations in peer reviewed health science literature. *BMJ (Online)*, 348, g1585. http://dx.doi.org/10.1136/bmj.g1585

Brazzeal, B. (2011). Citations to Wikipedia in chemistry journals: A preliminary study. *Issues in Science and Technology Librarianship*, 67. doi: 10.5062/F4057CV7

Buchanan, A.L., & Herubel, J.-P.V.M. (1993). Comparing materials used in philosophy and political science dissertations: A technical note. *Behavioral & Social Science Librarian*, 12(2), 63–70.

Budd, J. (1986). A citation study of American literature: implications for collection management. *Collection management*, 8(2), 49-62.

Butler, L. (2008). ICT assessment: Moving beyond journal outputs. *Scientometrics*, 74(1), 39–55.

Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., & Thijs, B. (2006). Traces of prior art: An analysis of non-patent references found in patent documents. *Scientometrics*, 69(1), 3-20.

Chen, L. (2007). Several colleges push to ban Wikipedia as resource, Duke Chronicle, available at: http://dukechronicle.com/article/several-colleges-push-ban-wikipedia-resource (accessed 29 July 2011).

Chen, C. & Roth, C. (2012). The dynamics of referencing in Wikipedia *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration.* New York, NY: ACM Press.

Chung, Y. K., (1995). Characteristics of references in international classification systems literature. *Library Quarterly*, 65(2), 200–215.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Lawrence Erlbaum Associates.

Costas, R., Zahedi, Z. & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003-2019.

Criscuolo, P., & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy*, *37*(10), 1892–1908. http://doi.org/10.1016/j.respol.2008.07.011

Cronin, B., Snyder, H.W., Rosenbaum, H., Martinson, A. & Callahan, E. (1998) Invoked on the Web. *Journal of the Association for Information Science and Technology*, 49(14), 1319-1328.

Cronin, B. (2001). Bibliometrics and beyond: some thoughts on web-based citation analysis. *Journal of Information Science*, 27(1), 1-7.

Cronin, B. (2013). The evolving indicator space (iSpace). *Journal of the Association for Information Science and Technology*, 64(8), 1523-1525.

Cronin, B. & Sugimoto, C.R. (2014). *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact.* Cambridge, MA: MIT Press.

Dellavalle, R.P., Schilling, L.M., Rodriguez, M.A., Van de Sompel, H. & Bollen, J. (2007). Refining dermatology journal impact factors using PageRank. *Journal of the American Academy of Dermatology*, *57*(1), 116-119.

Dowdy, S., Wearden, S., & Chilko, D. (2011). *Statistics for research.* New York, NY: John Wiley & Sons.

Drott, M. C. (1995). Reexamining the role of conference papers in scholarly communication. *Journal of the American Society for Information Science, 46*(4), 299.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152.

Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., & Zalányi, L. (2013). Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics*, *95*(1), 225–242. http://doi.org/10.1007/s11192-012-0796-4

Eichorn, P. & Yankauer, A. (1987). Do authors check their references? A survey of accuracy of references in 3 public-health journals, American Journal of Public Health, 77, 1011-1012.

Fleming, L., King, C., & Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science*, *18*(6), 938–954. http://doi.org/10.1287/orsc.1070.0289

Galloway, L. M., Pease, J. L., & Rauh, A. E. (2013). Introduction to altmetrics for science, technology, engineering, and mathematics (STEM) librarians. *Science & Technology Libraries, 32*(4), 335-345.

Glänzel, W., Schlemmer, B., Schubert, A., & Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics*, 68(3), 457–473.

Garfield, E. (1972), Citation analysis as a tool in journal evaluation – journals can be ranked by frequency and impact of citations for science policy studies, *Science*, 178, 471-9.

Goodrum, A. A., McCain, K. W., Lawrence, S., & Giles, C. L. (2001). Scholarly publishing in the internet age: A citation analysis of computer science literature. *Information Processing & Management, 37*(5), 661-675.

Griscom, R. (1983). Periodical use in a university music library: A citation study of theses and dissertations submitted to the Indiana University School of Music from 1975–1980. *The Serial Librarian*, 7(3), 35–52.

Gross, P.L. & Gross, E.M. (1927). College libraries and chemical education. *Science*, 66(1713), 385-389.

Gruzd, A., Goertzen, M. & Mai, P. (2012). *Survey results highlights: Trends in scholarly communication and knowledge dissemination in the age of social media*. Halifax, NS, Canada: Social Media Lab.

Gu, F. & Widén-Wulff, G. (2011). Scholarly communication and possible changes in the context of social media: A Finnish case study. *The Electronic Library*, 29(6), 762-776.

Gunn, W. (2013). Social signals reflect academic impact: What it means when a scholar adds a paper to Mendeley. *Information Standards Quarterly, 25*(2), 33-39.

Halavais, A. & Lackaff, D. (2008). An analysis of topical coverage of Wikipedia, *Journal of Computer-Mediated Communication*, 13(2), 429-440.

Haustein, S. & Larivière, V. (2014). Mendeley as a source of readership by students and postdocs? Evaluating article usage by academic status. *Proceedings of the 35th International Association of Technological University Libraries (IATUL)*, 2-5 June 2014, Helsinki, Finland. Available from: http://docs.lib.purdue.edu/iatul/2014/altmetrics/2

Haustein, S., & Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. J*ournal of Informetrics*, 5(3), 446-457.

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H. & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 101(2), 1145-1163.

Heck, T., Peters, I. & Stock, W.G. (2011). Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation. In *Proceedings of the 3rd ACM RecSys' 11 Workshop on Recommender Systems and*

the Social Web. New York, NY: ACM Press (pp. 16-23).
Heilman, J. M., Kemmann, E., Bonert, M., Chatterjee, A., Ragar, B., Beards, G. M., & Laurent, M. R. (2011). Wikipedia: A key tool for global public health promotion. *Journal of Medical Internet Research*, *13*(1). http://doi.org/10.2196/jmir.1589
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United states of America*, 102(46), 16569.
Holmberg, K. J. (2015). *Altmetrics for information professionals: Past, present and future*. London, UK: Chandos Publishing.
Huang, M., Huang, W., & Chen, D. (2014). Technological impact factor: An indicator to measure the impact of academic publications on practical innovation. *Journal of Informetrics*, 8(1), 241-251.
Huang, M. & Chang, Y. (2008). Characteristics of research output in social sciences and humanities: from a research evaluation perspective. *Journal of the American Society for Information Science and Technology*, 59(11), 1819–1828.
Ingwersen, P. (1998). The calculation of web impact factors. *Journal of documentation*, *54*(2), 236-243.
Jaschik, S. (2007). A stand against Wikipedia, Inside Higher Ed., available at: www.insidehighered.com/news/2007/01/26/wiki (accessed 29 July 2011).
Jeng, W., He, D. & Jiang, J. (2015). User participation in an academic social networking service: A survey of open group users on Mendeley. *Journal of the Association for Information Science and Technology*, 66(5), 890-904.
Jones, E., Kanouse, D., Kelley, H. H., Nisbett. R. E., Valins, S., & Weiner, B. (1972). *Attribution: Perceiving the causes of behavior*. Morristown, New Jersey: General Learning Press.
Jullien, N. (2012). What we know about Wikipedia: A review of the literature analyzing the project(s). https://hal.archives-ouvertes.fr/hal-00857208/document
Kademani, B., Anil, S. & Vijai, K. (1999). Conference papers of BARC scientists and engineers: A citation based study. In *National Conference on Putting Knowledge To Work*, Mumbai, India. http://eprints.rclis.org/13113/
Kapoor, R., Karvonen, M., Mohan, A. & Kässi, T. (2016). Patent citations as determinants of grant and opposition: case of European wind power industry. *Technology Analysis & Strategic Management*, 28(8), 950-964.
Katz, J. S. (2006). Introduction to a special issue on web indicators for innovation systems. *Research Evaluation*, *15*(2), 83–83. http://doi.org/10.3152/147154406781775977
Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *Algorithmica*, *1*(2), 1–9.
Knight, C., & Pryke, S. (2012). Wikipedia and the university, a case study. *Teaching in Higher Education*, 17(6), 649-659.
Knievel, J.E. & Kellsey, C. (2005). Citation analysis for collection development: A comparative study of eight humanities fields. *The Library Quarterly*, 75(2), 142-168.
Kousha, K. & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055-1065.
Kousha, K. & Thelwall, M. (2015a). An Automatic Method for Extracting Citations from Google Books. *Journal of the American Society for Information Science and Technology*, 66(2), 309–320.
Kousha, K. & Thelwall, M. (2015b). Can Amazon.com reviews help to assess the wider impacts of books? *Journal of the Association for Information Science and Technology*,

67(3), 566-581.
Kousha, K., & Thelwall, M. (2015c). Patent citation analysis with Google, *Journal of the Association for Information Science and Technology*, 68(1), 48-61. doi:10.1002/asi.23608.
Kousha, K., & Thelwall, M. (2009). Google book search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science & Technology*, *60*(8), 1537–1549.
Kousha, K., Thelwall, M. & Rezaie, S. (2011). Assessing the citation impact of books: The Role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology*, 62(11), 2147–2164.
Kousha, K., Thelwall, M. & Abdoli, M. (2016). Goodreads reviews to assess the wider impacts of books. *Journal of the Association for Information Science and Technology*, 68(8), 2004-2016.
Kousha, K. & Thelwall, M. (2017). Are Wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology*, 68(3), 762-779.
Kraker, P., Schlögl, C., Jack, K. and Lindstaedt, S. (2015). Visualization of co-readership patterns from an online reference management system. *Journal of Informetrics*, *9*(1), 169-182.
Lanjouw, J.O. & Schankerman, M. (2004). Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal*, 114(495), 441-465.
Laurent, M. R., & Vickers, T. J. (2009). Seeking Health Information Online: Does Wikipedia Matter? *Journal of the American Medical Informatics Association*, *16*(4), 471–479. http://doi.org/10.1197/jamia.M3059
Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411, 521-522.
Lee, C.J., Sugimoto, C.R., Zhang, G. & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2-17
Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. *Proceedings of the 17th International Conference on Science and Technology Indicators*, Vol. 2, (pp. 451-551).
Li, X., Thelwall, M. & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91(2), 461-471.
Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2003). The relationship between the WIFs or inlinks of Computer Science Departments in UK and their RAE ratings or research productivities in 2001. *Scientometrics*, 57(2), 239-255.
Liaw, Y., Chan, T., Fan, C., & Chiang, C. (2014). Can the technological impact of academic journals be evaluated? The practice of non-patent reference (NPR) analysis. *Scientometrics*, 101(1), 17-37. DOI: 10.1007/s11192-014-1337-0
Lisée, C., Larivière, V. & Archambault, É. (2008). Conference proceedings as a source of scientific information: A bibliometric analysis. *Journal of the Association for Information Science and Technology*, 59(11), 1776-1784.
Liu, J. & Adie, E. (2013). New perspectives on article-level metrics: developing ways to assess research uptake and impact online. *Insights*, 26(2), 153–158.
Lopez, P. (2010). Automatic extraction and resolution of bibliographical references in patent documents. Paper presented at the *Information Retrieval Facility Conference*.
MacRoberts, M.H. & MacRoberts, B.R. (1996). Problems of citation analysis. *Scientometrics***,** 36(3), 435-444.
MacRoberts, M.H., & MacRoberts, B.R. (2010). Problems of citation analysis: A study of

uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1-13.

Maflahi, N. & Thelwall, M. (2018). How quickly do publications get read? The evolution of Mendeley reader counts for new articles. *Journal of the Association for Information Science and Technology*, 69(1), 158-167.

Maflahi, N., & Thelwall, M. (2016). When are readership counts as useful as citation counts? Scopus vs. Mendeley for LIS journals. *Journal of the Association for information Science and Technology*, 67(1), 191-199.

Marcus, A. & Oransky, I. (2011). Science publishing: The paper is not sacred. *Nature*, 480(7378), 449-450.

Mas-Bleda, A. & Thelwall, M. (2016). Can alternative indicators overcome language biases in citation counts? A comparison of Spanish and UK research. *Scientometrics*, 109(3), 2007-2030.

Mas-Bleda, A., Thelwall, M., Kousha, K. & Aguillo, I.F. (2014). Do highly cited researchers successfully use the social web? *Scientometrics*, 101(1), 337-356.

McKerlich, R., Ives, C., & McGreal, R. (2013). Measuring use and creation of open educational resources in higher education. *International Review of Research in Open and Distance Learning*, *14*(4), 90–103.

Medelyan, O., Milne, D., Legg, C., & Witten, I. H. (2009). Mining meaning from Wikipedia. *International Journal of Human Computer Studies*, *67*(9), 716–754.

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F. Å., & Lanamäki, A. (2015). The sum of all human knowledge: A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology*, *66*(2),

Meyer, M., Debackere, K., & Glänzel, W. (2010). Can applied science be 'good science'? Exploring the relationship between patent citations and citation impact in nanoscience. *Scientometrics*, 85(2), 527-539.

Moed, H. F. (2009). New developments in the use of citation analysis in research evaluation. *Archivum Immunologiae et Therapiae Experimentalis*, *57*(1), 13–18. http://doi.org/10.1007/s00005-009-0001-5

Moed, H.F., Burger, W., Frankfort, J. & Van Raan, A.F. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, 14(3), 131-149.

Moed, H. (2005). *Citation analysis in research evaluation*. New York, NY: Springer.

Mohammadi, E. & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, 65(8), 1627-1638.

Mohammadi, E., Thelwall, M. & Kousha, K. (2013). Can Mendeley bookmarks reflect readership? A survey of user motivations. *Journal of the Association for Information Science and Technology*, 67(5), 1198-1209.

Mohammadi, E., Thelwall, M., Haustein, S. & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*, 66(9), 1832-1846.

Moskovkin, V., Shigorina, N. & Popov, D. (2012). The possibility of using the Google Patents search tool in patentometric analysis (based on the example of the world's largest innovative companies). *Scientific and Technical Information Processing*, 39(2), 107-112.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. *Research Policy*, *31*(8–9), 1389–1403. http://doi.org/10.1016/S0048-7333(02)00070-7

Narin, F. (1994). Patent bibliometrics. *Scientometrics*, *30*(1), 147–155. http://doi.org/10.1007/BF02017219

Narin, F., Carpenter, M. P., & Woolf, A. N. D. P. (1984). Technological performance assessments based on patents and patent citations, *IEEE Transactions on Engineering Management*, 1(4), 172-183.

Nederhof, A.J., & Noyons, E.C.M. (1992). International comparison of department's research performance in the humanities. *Journal of the American Society for Information Science*, 43(3), 249–256.

Nicolaisen, J. (2002), The J-shaped distribution of citedness, *Journal of Documentation*, 58(3), 383-395.

Nielsen, F.Å. (2008). Clustering of scientific citations in Wikipedia. *arXiv preprint arXiv:0805.1154*.

Orduna-Malea, E., Thelwall, M., & Kousha, K. (2017). Web citations in patents: Evidence of technological impact? *Journal of the Association for Information Science and Technology*, 68(8), 1967-1974. doi:10.1002/asi.23821

Park, T.K. (2011). The visibility of Wikipedia in scholarly publications, First Monday, 16(8), available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3492/3031 (accessed 08 October 2017).

Ponte, D. & Simon, J. (2011). Scholarly communication 2.0: Exploring researchers' opinions on Web 2.0 for scientific knowledge creation, evaluation and dissemination. *Serials Review*, *37*(3), 149-156.

Patterson, D. A. (2004). The health of research conferences and the dearth of big idea papers. *Communications of the ACM*, *47*(12), 23. http://doi.org/10.1145/1035134.1035153

Priem, J. (2014). Altmetrics. In: Cronin, B. & Sugimoto, C.R. (eds). *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact.* MIT Press. (pp.263-88).

Priem, J. & Hemminger, B.H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*, 15(7),

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. *ArXiv Preprint arXiv:1203.4745*,

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. http://altmetrics.org/manifesto/

Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A. & Asgari-Targhi, M. (2010). Adoption and use of Web 2.0 in scholarly communications. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4039-4056.

Rousseau, R. & Fred, Y.Y. (2013). A multi-metric approach for research evaluation. *Chinese Science Bulletin*, 58(26), 3288-3290.

Rowlands, I., Nicholas, D., Russell, B., Canty, N. & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, 24(3), 183-195.

Royal, C. & Kapila, C. (2008). What's on Wikipedia, and what's not.? Assessing completeness of information, *Social Science Computer Review*, 27(1), 138-148.

Sandstrom, U., Wadskog, D. & Karlsson, S. (2005). Research institutes and universities: does collaboration pay?, in Ingwersen, P. and Larsen, B. (Eds), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics*, Karolinska University Press, Stockholm.

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K. & Kraker, P. (2014). Comparison of downloads, citations and readership data for two information systems journals. *Scientometrics*, 101(2), 1113-1128.

Schmoch, U. (1993). Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics*, 26(1), 193-211.

Shema, H., Bar-Ilan, J. & Thelwall, M. (2012). Research blogs and the discussion of scholarly information. *PloS One*, 7(5), e35869.

Shema, H., Bar-Ilan, J. & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for information science and technology*, 65(5), 1018-1027.

Shirabe, M. (2014). Identifying SCI covered publications within non-patent references in US utility patents. *Scientometrics*, *101*(2), 999-1014.

Smart, S. & Waldfogel, J. (1996). *A citation-based test for discrimination at economics and finance journals*, NBER Working Paper No. 5460, National Bureau of Economic Research, Cambridge, MA.

Smith, D.R. (2012). Impact factors, scientometrics and the history of citation-based research. *Scientometrics*, 92(2), 419-427.

Smith, A. & Thelwall, M. (2002). Web impact factors for Australasian universities. *Scientometrics*, 54(3), 363-380.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*(7), 994–1017. http://doi.org/10.1016/j.respol.2006.05.002

Soules, A. (2015). Faculty perception of Wikipedia in the California state university system. *New Library World*, 116(3-4), 213-226.

Stuart, D. & Thelwall, M. (2005). What can university-to-government web links reveal about university-government collaborations *Proceedings of the 10th International Conference of the International Society of Scientometrics and Informetrics.*(pp.188-192).

Sud, P. & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98(2), 1131-1143. 10.1007/s11192-013-1117-2.

Suh, B., Chi, H.E., Kittur, A. & Pendleton, B.A. (2008). Lifting the Veil: improving account-ability and social transparency in Wikipedia with Wiki Dashboard, CHI '08 *Proceeding of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, 1037-1040.

Tang, R. & Thelwall, M. (2004). US academic departmental Web-site interlinking in the United States disciplinary differences. *Library & Information Science Research*, 25(4), 437-458.

Terry, D. (2014). Publish now, judge later. *Communications of the ACM*, 57(1), 44-46.

Thelwall, M. (in press). Do Mendeley reader counts indicate the value of arts and humanities research? *Journal of Librarianship and Information Science*, 0961000617732381.

Thelwall, M. (2012). Journal impact evaluation: a webometric perspective. *Scientometrics*, 92(2), 429-441.

Thelwall, M. & Sud, P. (2015). Mendeley readership counts: An investigation of temporal and disciplinary differences. *Journal of the Association for Information Science and Technology*, 57(6), 3036-3050.

Thelwall, M. (2016b). Interpreting correlations between citation counts and other indicators. *Scientometrics*, 108(1), 337-347.

Thelwall, M. (2017a). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics*, 11(1), 128–151. 10.1016/j.joi.2016.12.002

Thelwall, M. (2017b). Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa? *Journal of Librarianship and Information Science*, 49(2), 144-151.

Thelwall, M. (2017c). Are Mendeley reader counts high enough for research evaluations when articles are published? *Aslib Journal of Information Management*, 69(2), 2-34.

Thelwall, M. (2017d). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics*, 113(3), 1721–1731.
Thelwall, M., & Kousha, K. (2008). Online presentations as a source of scientific impact? An analysis of PowerPoint files citing academic journals. *Journal of the American Society for Information Science and Technology*, 59(5), 805-815.
Thelwall, M., & Wilson, P. (2014a). Regression for citation data: An evaluation of different methods. *Journal of Informetrics*, 8(4), 963-971.
Thelwall, M. & Wilson, P. (2014b). Distributions for cited articles from individual subjects and years. *Journal of Informetrics*, 8(4), 824-83
Thelwall, M. & Wilson, P. (2016). Mendeley readership altmetrics for medical articles: An analysis of 45 fields. *Journal of the Association for Information Science and Technology*, 67(8), 1962-1972.
Thelwall, M., & Kousha, K. (2015). Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. *El Profesional de la Información*, 24(5), 587-606.
Thelwall, M. & Maflahi, N. (2015). How important is computing technology for library and information science research? *Library & Information Science Research*, 37(1), 42-50.
Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? twitter and ten other social web services. *PloS One*, *8*(5), e64841.
Thomas, G. R., Eng, L., de Wolff, J. F., & Grover, S. C. (2013). An Evaluation of Wikipedia as a Resource for Patient Education in Nephrology. *Seminars in Dialysis*, *26*(2), 159–163. http://doi.org/10.1111/sdi.12059
Tomaszewski, R., & MacDonald, K.I. (2016). A Study of Citations to Wikipedia in Scholarly Publications. *Science & Technology Libraries*, 35(3), 246-261.
Torres-Salinas, D., Cabezas-Clavijo, Á., & Jiménez-Contreras, E. (2013). Altmetrics: New indicators for scientific communication in web 2.0. *ArXiv Preprint arXiv:1306.659.*
Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, 21(1), 172-187.
Tussen, R., Buter, R., & Van Leeuwen, T. (2000). Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics*, 47(2), 389-412.
Van Raan, A.F. (2004). Sleeping beauties in science. *Scientometrics*, 59(3), 467-472.
Van Raan, A.F.J. (2005). For your citations only? Hot topics in bibliometric analysis, *Measurement: Interdisciplinary Research and Perspectives*, 3(1), 50-62.
Vardi, M. Y. (2009). Conferences vs. journals in computing research. *Communications of the ACM*, *52*(5), 5. http://doi.org/10.1145/1506409.1506410
Vaughan, L. & Hysen, K. (2002). Relationship between links to journal Web sites and impact factors. *Aslib Proceedings*, 24(3), 356-361.
Vaughan, L. & Romero-Frías, E. (2014). Web search volume as a predictor of academic fame: an exploration of Google Trends. *Journal of the Association for Information Science and Technology*, 65(4), 707-720.
Vaughan, L. & Shaw, D. (2003). Bibliographic and web citations: what is the difference? *Journal of the Association for Information Science and Technology*, 54(14), 1313-1322.
Vaughan, L. & Shaw, D. (2005). Web citation data for impact assessment: A comparison of four science disciplines. *Journal of the Association for Information Science and Technology*, 56(10), 1075-1087.
Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics*, *54*(3), 399-420.

Visser, M.S., & Moed, H.F. (2005). Developing bibliometric indicators of research perfor mance in computer science. In P. Ingwersen & B. Larsen (Eds.), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (275–279). Stockholm: Karolinska University Press.

Vrettas, G. & Sanderson, M. (2015). Conferences versus journals in computer science. *Journal of the Association for Information Science and Technology*, 66(12), 2674-2684.

Waltman, Ludo, Van Eck, N. J., Van Leeuwen, T. N., Visser, M. S., & Van Raan, A. F. J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37–47.

Wang, X., Zhang, X., & Xu, S. (2011). Patent co-citation networks of Fortune 500 companies. *Scientometrics*, *88*(3), 761–770. http://doi.org/10.1007/s11192-011-0414-x

Waters, N.L. (2007). Why you can't cite Wikipedia in my class, *Communications of the ACM*, 50(9), available at: www.inf.ed.ac.uk/teaching/courses/irm/p15-waters.pdf (accessed 08 November 2017).

White, M. J. (2015). Patent Sightings: A comparative analysis of patent citation search tools using case studies from the engineering literature. *2015 122nd ASEE Annual Conference and Exposition, 122nd ASEE* (122nd ASEE Annual Conference and Exposition: Making Value for Society.

Wilkinson, D. & Huberman, B. (2007). Cooperation and quality in Wikipedia, WikiSym *Proceedings of the 2007 international symposium on Wikis*, ACM, New York, NY. (pp. 157-164).

Wilsdon, J. (2016). *The metric tide: independent review of the role of metrics in research assessment and management.* Oxford, UK: Sage.

Wouters, P., & Costas, R. (2012). *Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century* SURFfoundation Utrecht.

Yasseri, T., & Kertész, J. (2013). Value production in a collaborative environment. *Journal of Statistical Physics*, *151*(3-4), 414–439.

Zahedi, Z., Costas, R., Larivière, V. & Haustein, S. (2017). What makes papers visible on social media? An analysis of various document characteristics. *arXiv preprint arXiv:1703.05777*.

Zahedi, Z., Costas, R. & Wouters, P. (2015). Do Mendeley readership counts help to filter highly cited WoS publications better than average citation impact of journals (JCS)? *arXiv preprint arXiv:1507.02093*.

Zahedi, Z., Costas, R. & Wouters, P. (2014a). Broad altmetric analysis of Mendeley readerships through the 'academic status' of the readers of scientific publications. *STI 2014*, Leiden, NL: Universiteit Leiden *(pp.* 720-727).

Zahedi, Z., Costas, R., & Wouters, P. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? *Plos ALM workshop*, San Francisco.

Zahedi, Z., Costas, R., & Wouters, P. (2014b). How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics, 101*(2), 1491-1513.

Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. *TechTrends*, *55*(1), 32–36. http://doi.org/10.1007/s11528-011-0467-y

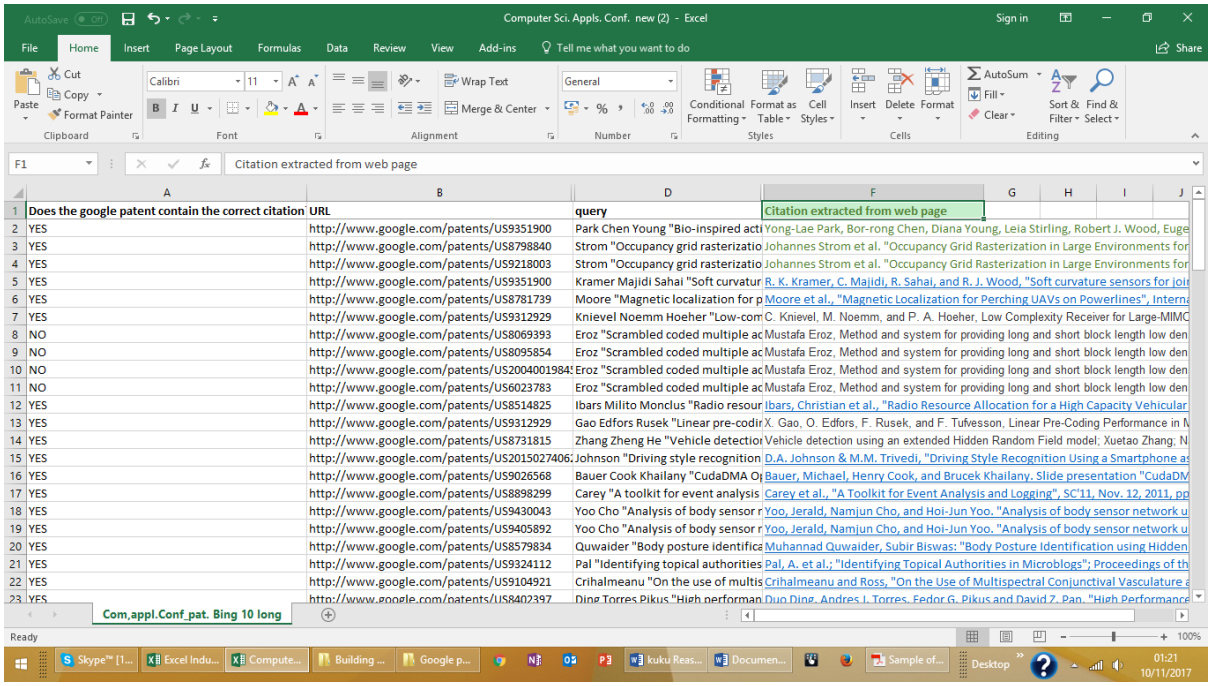
Zhang, Y. H., & Jia, X. (2013). Republication of conference papers in journals? *Learned Publishing*, *26*(3), 189–196. http://doi.org/10.1087/20130307

Ziman, J. (2000). *Real Science. What it is, and what it means*. Cambridge University Press, Cambridge.

Zimmermann, C. (2013). Academic rankings with RePEc. *Econometrics*, 1(3), 249-280.

Appendices

Appendix 1 Screenshot of Google Patents manual checks for Computer Science Applications



Computer Sci. Appls. Conf. new (2) - Excel

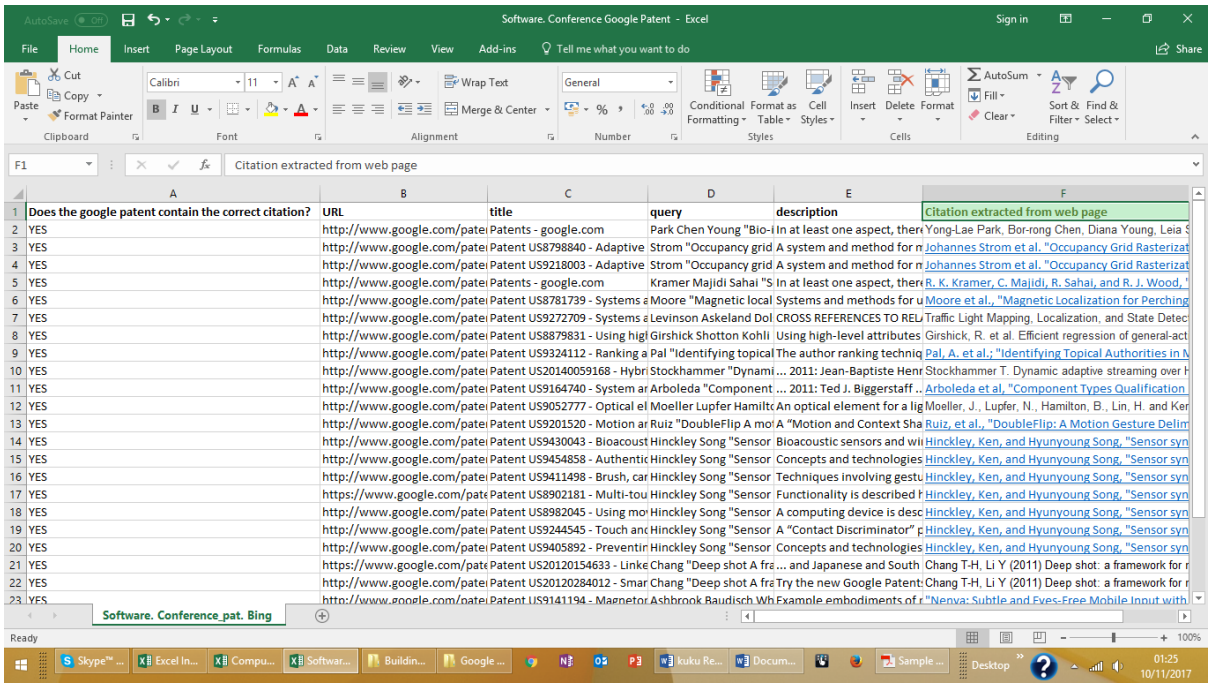
Citation extracted from web page

	A	B	D	F
1	Does the google patent contain the correct citation	URL	query	Citation extracted from web page
2	YES	http://www.google.com/patents/US9351900	Park Chen Young "Bio-inspired acti	Yong-Lae Park, Bor-rong Chen, Diana Young, Leia Stirling, Robert J. Wood, Euge
3	YES	http://www.google.com/patents/US8798840	Strom "Occupancy grid rasterizatio	Johannes Strom et al. "Occupancy Grid Rasterization in Large Environments for
4	YES	http://www.google.com/patents/US9218003	Strom "Occupancy grid rasterizatio	Johannes Strom et al. "Occupancy Grid Rasterization in Large Environments for
5	YES	http://www.google.com/patents/US9351900	Kramer Majidi Sahai "Soft curvatur	R. K. Kramer, C. Majidi, R. Sahai, and R. J. Wood, "Soft curvature sensors for joi
6	YES	http://www.google.com/patents/US8781739	Moore "Magnetic localization for p	Moore et al., "Magnetic Localization for Perching UAVs on Powerlines", Interna
7	YES	http://www.google.com/patents/US9312929	Knievel Noemm Hoeher "Low-com	C. Knievel, M. Noemm, and P. A. Hoeher, Low Complexity Receiver for Large-MIMO
8	NO	http://www.google.com/patents/US8069393	Eroz "Scrambled coded multiple ac	Mustafa Eroz, Method and system for providing long and short block length low den
9	NO	http://www.google.com/patents/US8095854	Eroz "Scrambled coded multiple ac	Mustafa Eroz, Method and system for providing long and short block length low den
10	NO	http://www.google.com/patents/US20040019845	Eroz "Scrambled coded multiple ac	Mustafa Eroz, Method and system for providing long and short block length low den
11	NO	http://www.google.com/patents/US6023783	Eroz "Scrambled coded multiple ac	Mustafa Eroz, Method and system for providing long and short block length low den
12	YES	http://www.google.com/patents/US8514825	Ibars Milito Monclus "Radio resour	Ibars, Christian et al., "Radio Resource Allocation for a High Capacity Vehicular
13	YES	http://www.google.com/patents/US9312929	Gao Edfors Rusek "Linear pre-codir	X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, Linear Pre-Coding Performance in N
14	YES	http://www.google.com/patents/US8731815	Zhang Zheng He "Vehicle detectio	Vehicle detection using an extended Hidden Random Field model; Xuetao Zhang; N
15	YES	http://www.google.com/patents/US20150274065	Johnson "Driving style recognition	D.A. Johnson & M.M. Trivedi, "Driving Style Recognition Using a Smartphone as
16	YES	http://www.google.com/patents/US9026568	Bauer Cook Khailany "CudaDMA Op	Bauer, Michael, Henry Cook, and Brucek Khailany. Slide presentation "CudaDM
17	YES	http://www.google.com/patents/US8898299	Carey "A toolkit for event analysis	Carey et al., "A Toolkit for Event Analysis and Logging", SC'11, Nov. 12, 2011, pp
18	YES	http://www.google.com/patents/US9430043	Yoo Cho "Analysis of body sensor r	Yoo, Jerald, Namjun Cho, and Hoi-Jun Yoo. "Analysis of body sensor network u
19	YES	http://www.google.com/patents/US9405892	Yoo Cho "Analysis of body sensor r	Yoo, Jerald, Namjun Cho, and Hoi-Jun Yoo. "Analysis of body sensor network u
20	YES	http://www.google.com/patents/US8579834	Quwaider "Body posture identifica	Muhannad Quwaider, Subir Biswas: "Body Posture Identification using Hidden
21	YES	http://www.google.com/patents/US9324112	Pal "Identifying topical authorities	Pal, A. et al.; "Identifying Topical Authorities in Microblogs"; Proceedings of th
22	YES	http://www.google.com/patents/US9104921	Crihalmeanu "On the use of multis	Crihalmeanu and Ross, "On the Use of Multispectral Conjunctival Vasculature a
23	YES	http://www.google.com/patents/US8402397	Ding Torres Pikus "High performan	Duo Ding, Andres J. Torres, Fedor G. Pikus and David Z. Pan, "High Performance

Com,appl.Conf_pat. Bing 10 long

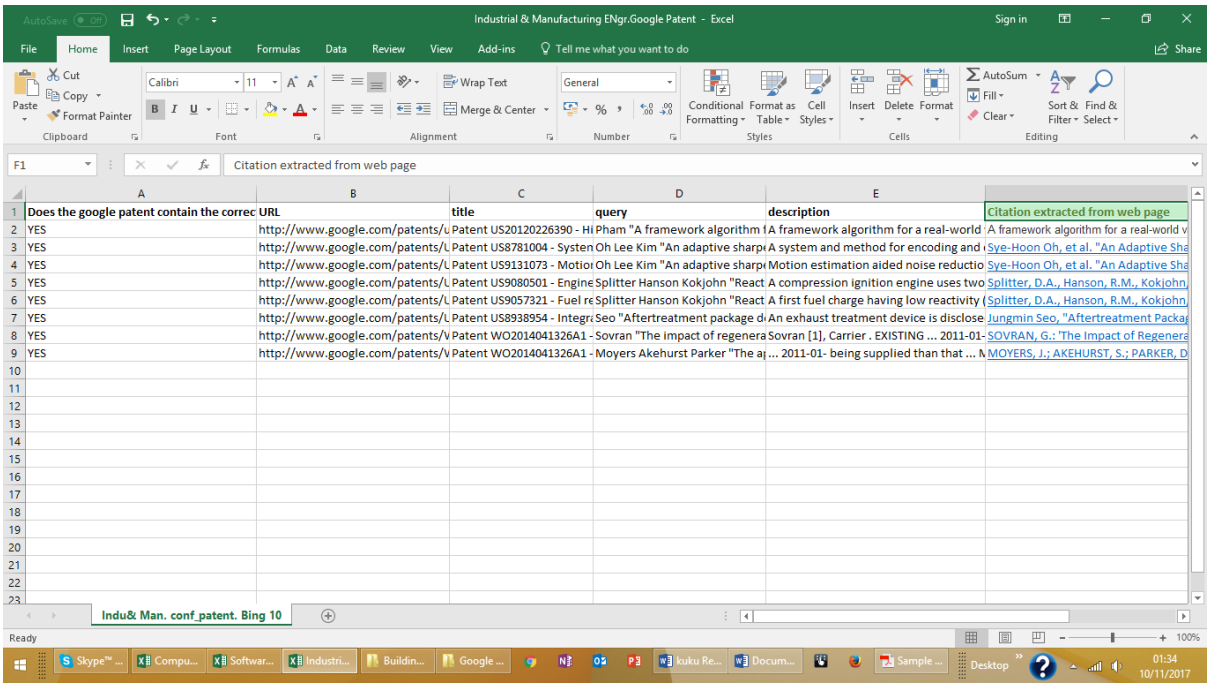
Ready

Appendix 2 Screenshot of Google Patents manual checks for Software Engineering



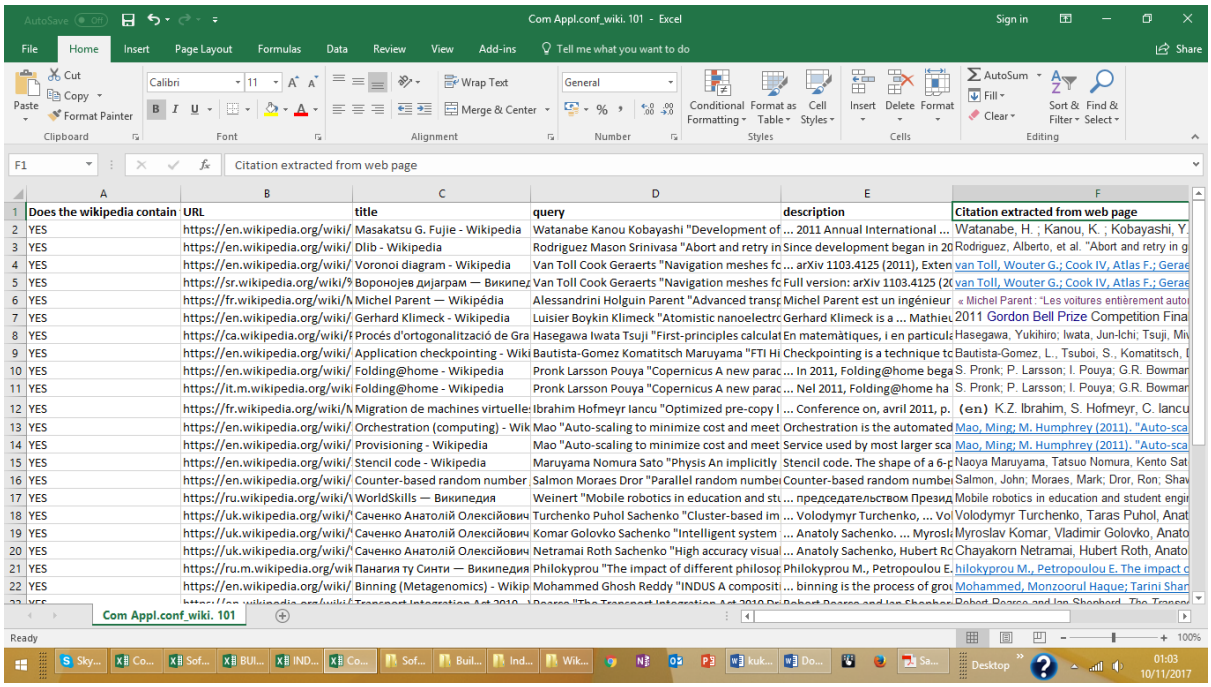
Does the google patent contain the correct citation?	URL	title	query	description	Citation extracted from web page
YES	http://www.google.com/patei	Patents - google.com	Park Chen Young "Bio-i	In at least one aspect, there	Yong-Lae Park, Bor-rong Chen, Diana Young, Leia S
YES	http://www.google.com/patei	Patent US8798840 - Adaptive	Strom "Occupancy grid	A system and method for n	Johannes Strom et al. "Occupancy Grid Rasterizat
YES	http://www.google.com/patei	Patent US9218003 - Adaptive	Strom "Occupancy grid	A system and method for n	Johannes Strom et al. "Occupancy Grid Rasterizat
YES	http://www.google.com/patei	Patents - google.com	Kramer Majidi Sahai "S	In at least one aspect, there	R. K. Kramer, C. Majidi, R. Sahai, and R. J. Wood, "
YES	http://www.google.com/patei	Patent US8781739 - Systems a	Moore "Magnetic local	Systems and methods for u	Moore et al., "Magnetic Localization for Perching
YES	http://www.google.com/patei	Patent US9272709 - Systems a	Levinson Askeland Dol	CROSS REFERENCES TO REL	Traffic Light Mapping, Localization, and State Detec
YES	http://www.google.com/patei	Patent US8879831 - Using high	Girshick Shotton Kohli	Using high-level attributes	Girshick, R. et al. Efficient regression of general-act
YES	http://www.google.com/patei	Patent US9324112 - Ranking a	Pal "Identifying topical	The author ranking techniq	Pal, A. et al.; "Identifying Topical Authorities in M
YES	http://www.google.com/patei	Patent US20140059168 - Hybri	Stockhammer "Dynami	... 2011: Jean-Baptiste Henr	Stockhammer T. Dynamic adaptive streaming over H
YES	http://www.google.com/patei	Patent US9164740 - System ar	Arboleda "Component	... 2011: Ted J. Biggerstaff ..	Arboleda et al, "Component Types Qualification
YES	http://www.google.com/patei	Patent US9052777 - Optical el	Moeller Lupfer Hamilto	An optical element for a lig	Moeller, J., Lupfer, N., Hamilton, B., Lin, H. and Ker
YES	http://www.google.com/patei	Patent US9201520 - Motion ar	Ruiz "DoubleFlip A mo	A "Motion and Context Sha	Ruiz, et al., "DoubleFlip: A Motion Gesture Delim
YES	http://www.google.com/patei	Patent US9430043 - Bioacoust	Hinckley Song "Sensor	Bioacoustic sensors and wi	Hinckley, Ken, and Hyunyoung Song, "Sensor syn
YES	http://www.google.com/patei	Patent US9454858 - Authentic	Hinckley Song "Sensor	Concepts and technologies	Hinckley, Ken, and Hyunyoung Song, "Sensor syn
YES	http://www.google.com/patei	Patent US9411498 - Brush, car	Hinckley Song "Sensor	Techniques involving gestu	Hinckley, Ken, and Hyunyoung Song, "Sensor syn
YES	https://www.google.com/pate	Patent US8902181 - Multi-tou	Hinckley Song "Sensor	Functionality is described h	Hinckley, Ken, and Hyunyoung Song, "Sensor syn
YES	http://www.google.com/patei	Patent US8982045 - Using mo	Hinckley Song "Sensor	A computing device is desc	Hinckley, Ken, and Hyunyoung Song, "Sensor syn
YES	http://www.google.com/patei	Patent US9244545 - Touch and	Hinckley Song "Sensor	A "Contact Discriminator" p	Hinckley, Ken, and Hyunyoung Song, "Sensor syn
YES	http://www.google.com/patei	Patent US9405892 - Preventin	Hinckley Song "Sensor	Concepts and technologies	Hinckley, Ken, and Hyunyoung Song, "Sensor syn
YES	https://www.google.com/pate	Patent US20120154633 - Linke	Chang "Deep shot A fra	... and Japanese and South	Chang T-H, Li Y (2011) Deep shot: a framework for r
YES	http://www.google.com/patei	Patent US20120284012 - Smar	Chang "Deep shot A fra	Try the new Google Patent:	Chang T-H, Li Y (2011) Deep shot: a framework for r
YES	http://www.google.com/patei	Patent US9141194 - Magnetor	Ashbrook Baudisch Wh	Example embodiments of r	"Nenya: Subtle and Eyes-Free Mobile Input with

Appendix 3 Screenshot of Google Patents manual checks for Industrial & Manufacturing Engineering

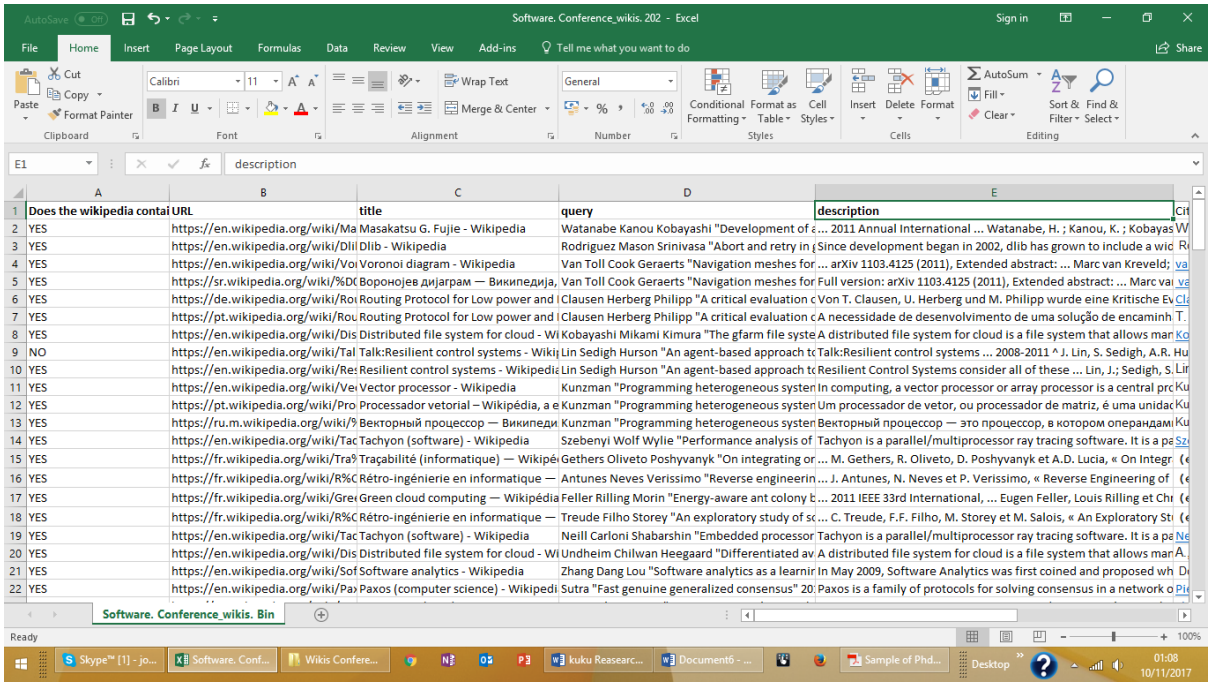


Does the google patent contain the correc	URL	title	query	description	Citation extracted from web page
YES	http://www.google.com/patents/u	Patent US20120226390 - Hi	Pham "A framework algorithm f	A framework algorithm for a real-world	A framework algorithm for a real-world v
YES	http://www.google.com/patents/U	Patent US8781004 - Systen	Oh Lee Kim "An adaptive sharpe	A system and method for encoding and	Sye-Hoon Oh, et al. "An Adaptive Sha
YES	http://www.google.com/patents/U	Patent US9131073 - Motior	Oh Lee Kim "An adaptive sharpe	Motion estimation aided noise reductio	Sye-Hoon Oh, et al. "An Adaptive Sha
YES	http://www.google.com/patents/U	Patent US9080501 - Engine	Splitter Hanson Kokjohn "React	A compression ignition engine uses two	Splitter, D.A., Hanson, R.M., Kokjohn,
YES	http://www.google.com/patents/U	Patent US9057321 - Fuel re	Splitter Hanson Kokjohn "React	A first fuel charge having low reactivity (	Splitter, D.A., Hanson, R.M., Kokjohn,
YES	http://www.google.com/patents/U	Patent US8938954 - Integra	Seo "Aftertreatment package d	An exhaust treatment device is disclose	Jungmin Seo, "Aftertreatment Packag
YES	http://www.google.com/patents/W	Patent WO2014041326A1 -	Sovran "The impact of regenera	Sovran [1], Carrier . EXISTING ... 2011-01-	SOVRAN, G.: 'The Impact of Regenera
YES	http://www.google.com/patents/W	Patent WO2014041326A1 -	Moyers Akehurst Parker "The ap	... 2011-01- being supplied than that ... M	MOYERS, J.; AKEHURST, S.; PARKER, D

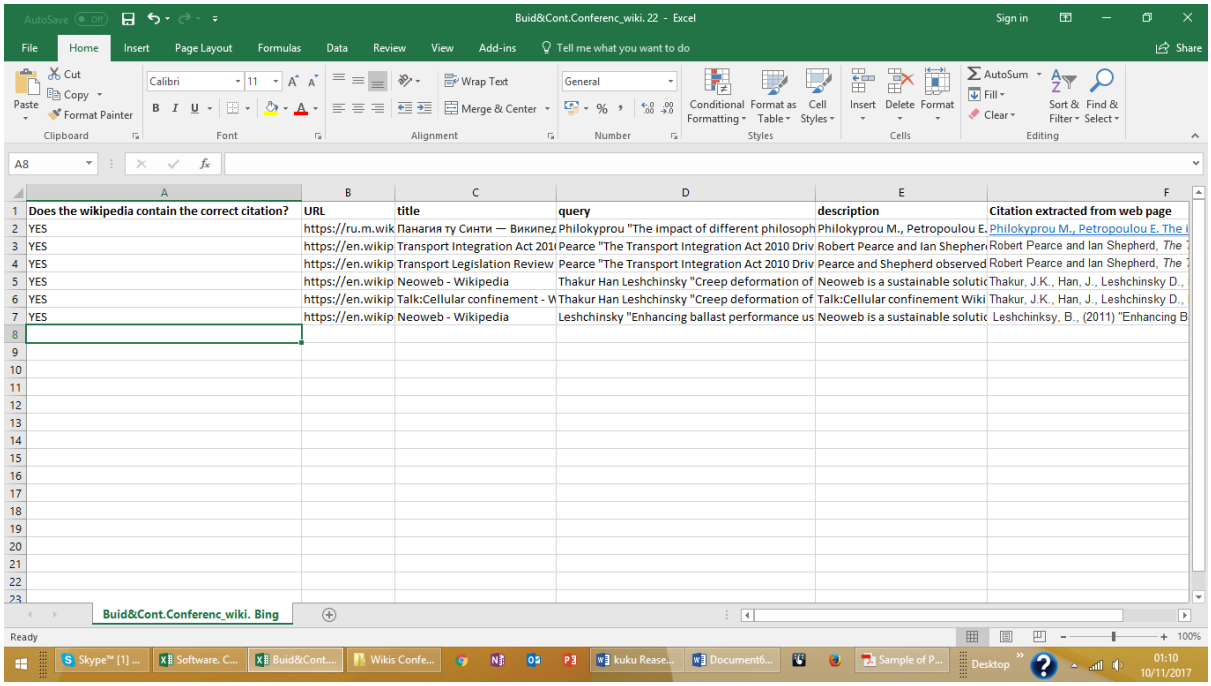
Appendix 4 Screenshot of Wikipedia citations manual checks for Computer Science Applications



Appendix 5 Screenshot of Wikipedia citations manual checks for Software Engineering

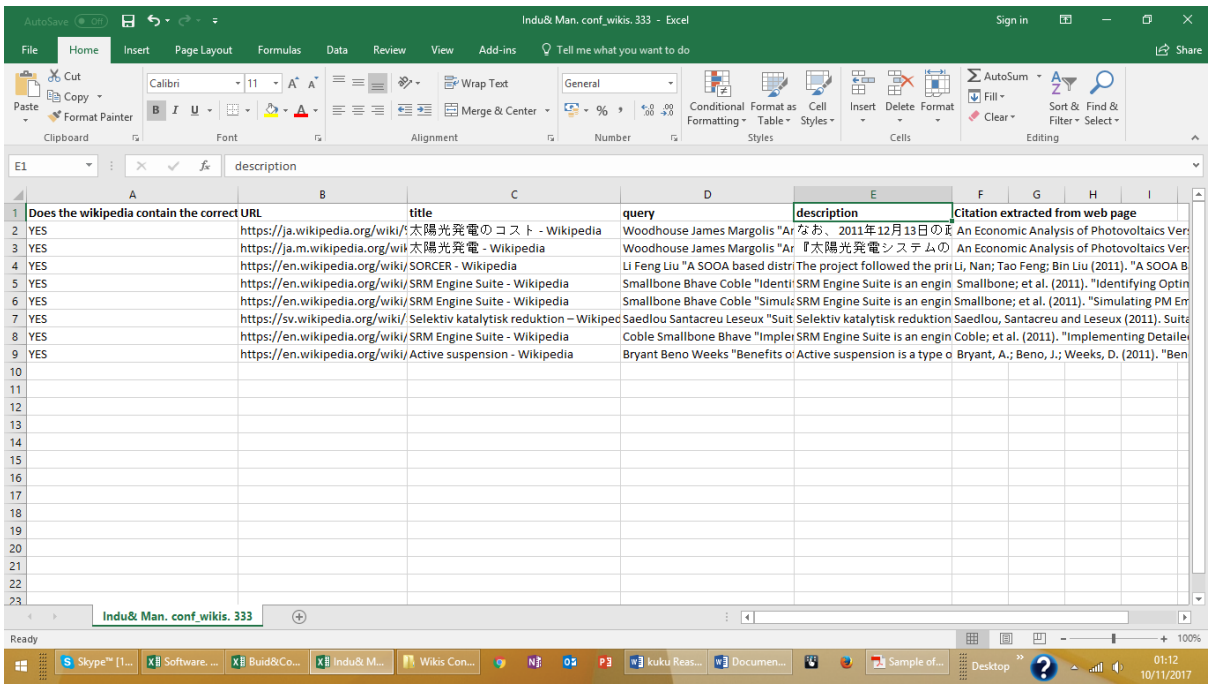


Appendix 6 Screenshot of Wikipedia citations manual checks for Building & Construction Engineering



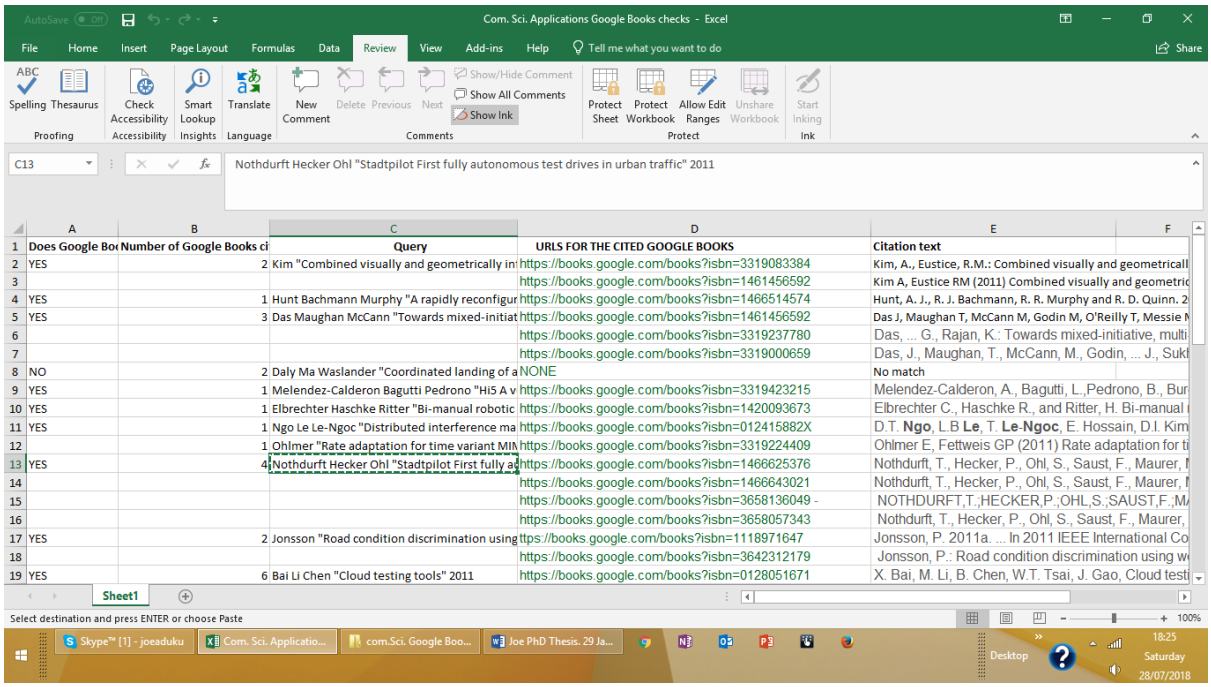
Does the wikipedia contain the correct citation?	URL	title	query	description	Citation extracted from web page
YES	https://ru.m.wik	Панагия ту Синти — Википед	Philokyprou "The impact of different philosoph	Philokyprou M., Petropoulou E.	Philokyprou M., Petropoulou E. The
YES	https://en.wikip	Transport Integration Act 201	Pearce "The Transport Integration Act 2010 Driv	Robert Pearce and Ian Shepher	Robert Pearce and Ian Shepherd, *The*
YES	https://en.wikip	Transport Legislation Review	Pearce "The Transport Integration Act 2010 Driv	Pearce and Shepherd observed	Robert Pearce and Ian Shepherd, *The*
YES	https://en.wikip	Neoweb - Wikipedia	Thakur Han Leshchinsky "Creep deformation of	Neoweb is a sustainable solutic	Thakur, J.K., Han, J., Leshchinsky D.,
YES	https://en.wikip	Talk:Cellular confinement - W	Thakur Han Leshchinsky "Creep deformation of	Talk:Cellular confinement Wiki	Thakur, J.K., Han, J., Leshchinsky D.,
YES	https://en.wikip	Neoweb - Wikipedia	Leshchinsky "Enhancing ballast performance us	Neoweb is a sustainable solutic	Leshchinksy, B., (2011) "Enhancing B

Appendix 7 Screenshot of Wikipedia citations manual checks for Industrial & Manufacturing Engineering

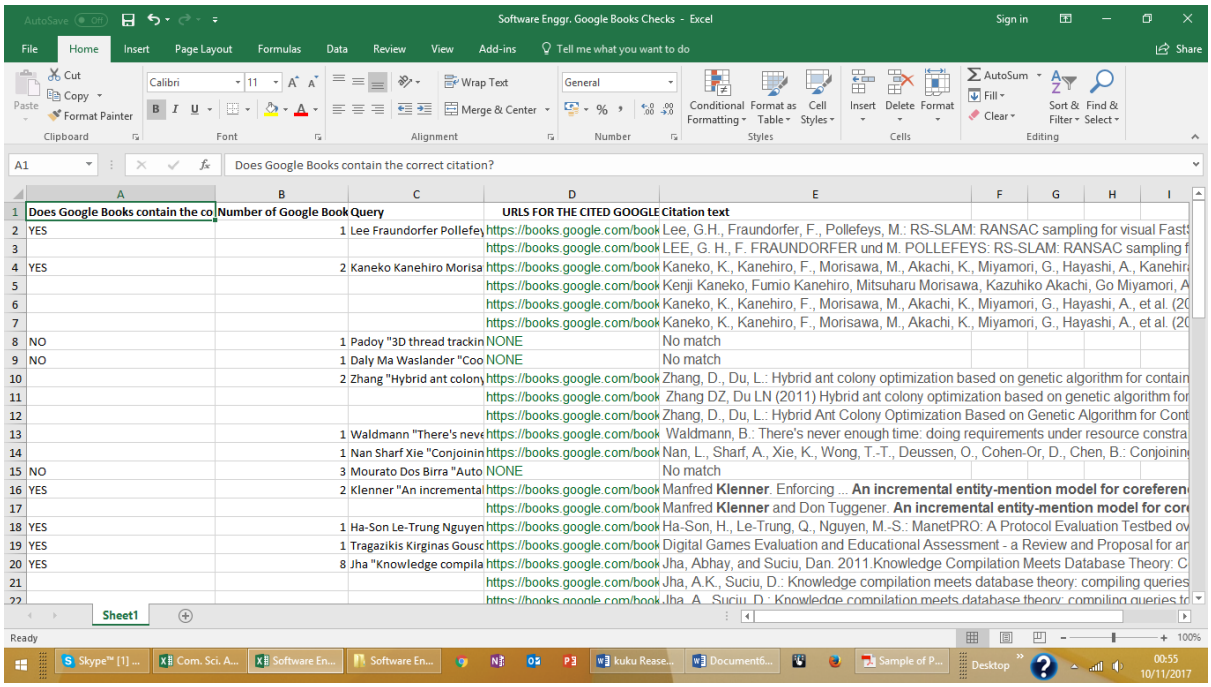


Does the wikipedia contain the correct	URL	title	query	description	Citation extracted from web page
YES	https://ja.wikipedia.org/wiki/	太陽光発電のコスト - Wikipedia	Woodhouse James Margolis "Ar	なお、2011年12月13日の	An Economic Analysis of Photovoltaics Ver
YES	https://ja.m.wikipedia.org/wik	太陽光発電 - Wikipedia	Woodhouse James Margolis "Ar	『太陽光発電システムの	An Economic Analysis of Photovoltaics Ver
YES	https://en.wikipedia.org/wiki/	SORCER - Wikipedia	Li Feng Liu "A SOOA based distr	The project followed the pri	Li, Nan; Tao Feng; Bin Liu (2011). "A SOOA B
YES	https://en.wikipedia.org/wiki/	SRM Engine Suite - Wikipedia	Smallbone Bhave Coble "Identi	SRM Engine Suite is an engin	Smallbone; et al. (2011). "Identifying Optin
YES	https://en.wikipedia.org/wiki/	SRM Engine Suite - Wikipedia	Smallbone Bhave Coble "Simul	SRM Engine Suite is an engin	Smallbone; et al. (2011). "Simulating PM Em
YES	https://sv.wikipedia.org/wiki/	Selektiv katalytisk reduktion – Wikiped	Saedlou Santacreu Leseux "Suit	Selektiv katalytisk reduktion	Saedlou, Santacreu and Leseux (2011). Suita
YES	https://en.wikipedia.org/wiki/	SRM Engine Suite - Wikipedia	Coble Smallbone Bhave "Imple	SRM Engine Suite is an engin	Coble; et al. (2011). "Implementing Detailed
YES	https://en.wikipedia.org/wiki/	Active suspension - Wikipedia	Bryant Beno Weeks "Benefits o	Active suspension is a type o	Bryant, A.; Beno, J.; Weeks, D. (2011). "Ben

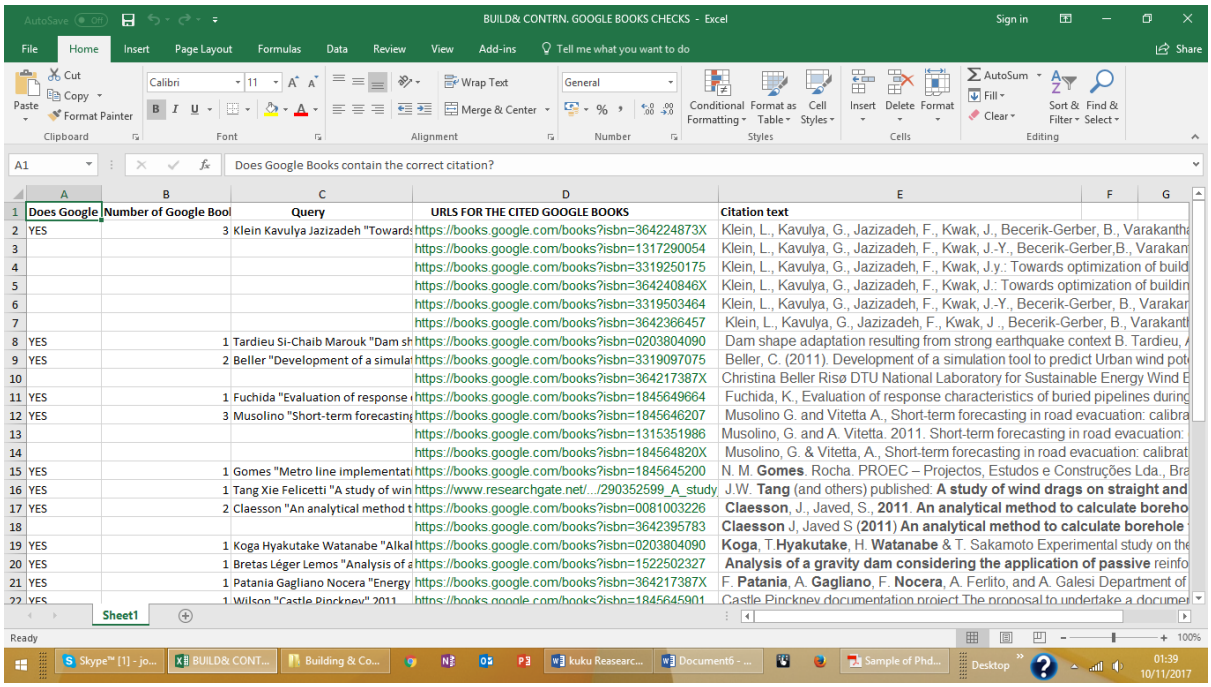
Appendix 8 Screenshot of Google Books citations manual checks for Computer Science Applications



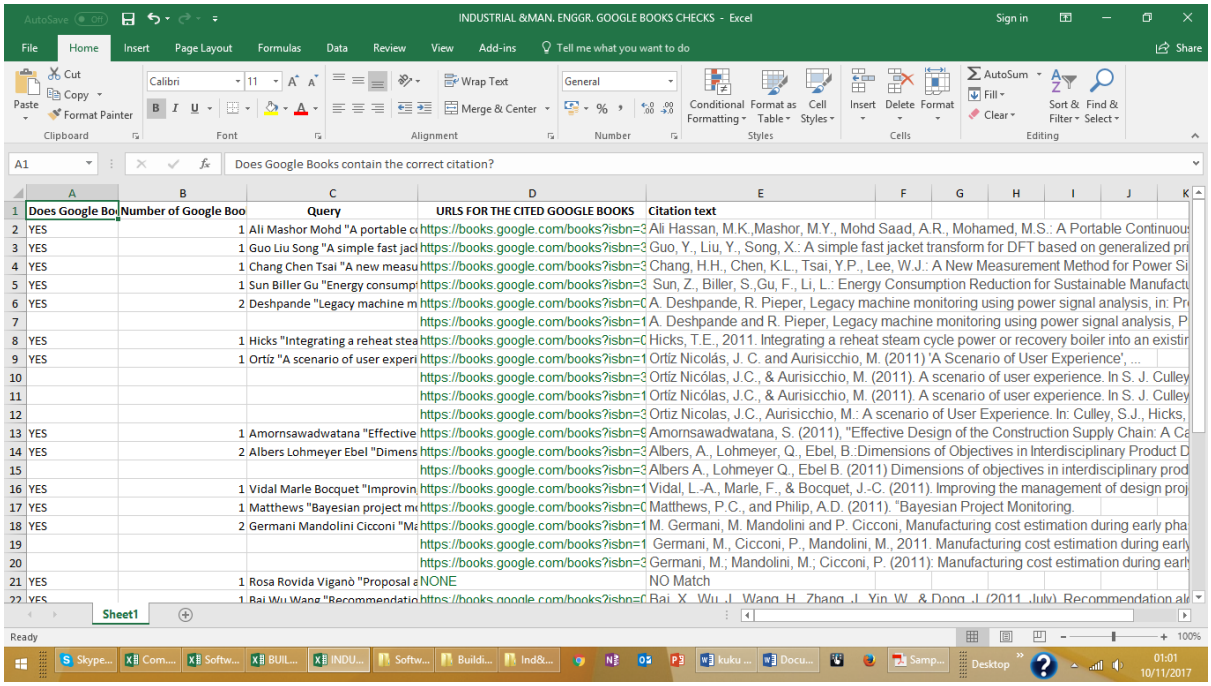
Appendix 9 Screenshot of Google Books citations manual checks for Software Engineering



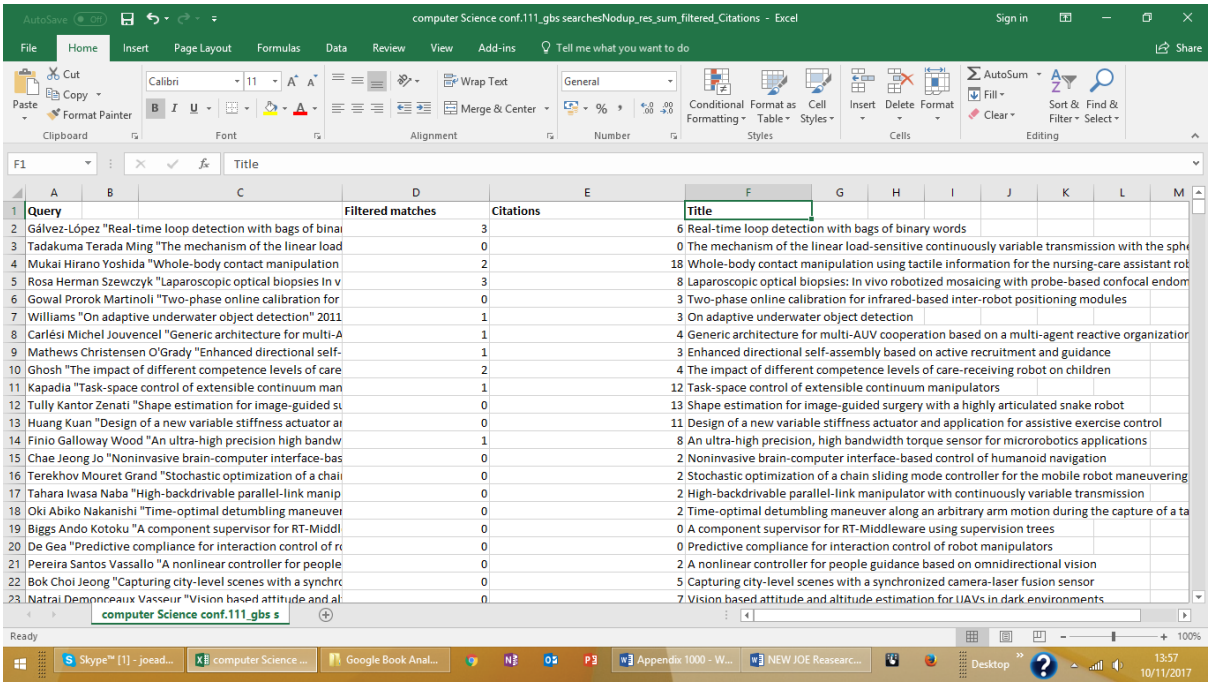
Appendix 10 Screenshot of Google Books citations manual checks for Building & Construction Engineering



Appendix 11 Screenshot of Google Books citations manual checks for Industrial & Manufacturing Engineering

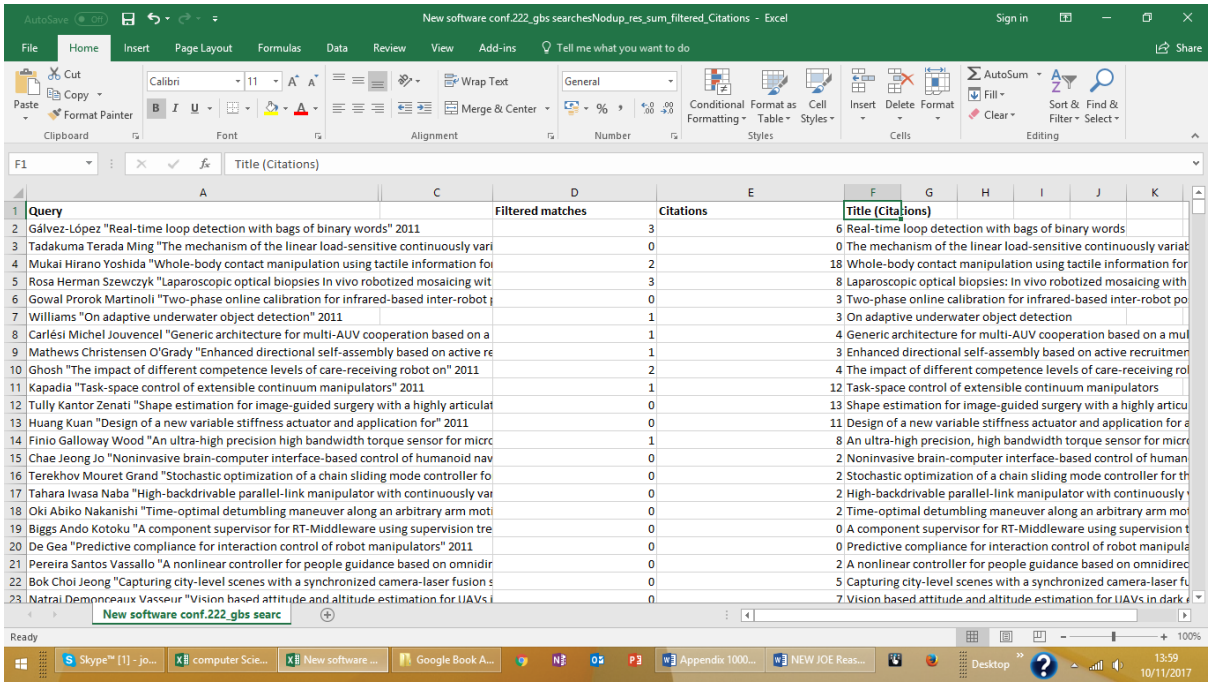


Appendix 12 Screenshot of Google Books filtered citation counts and Scopus citation counts for Computer Science Applications



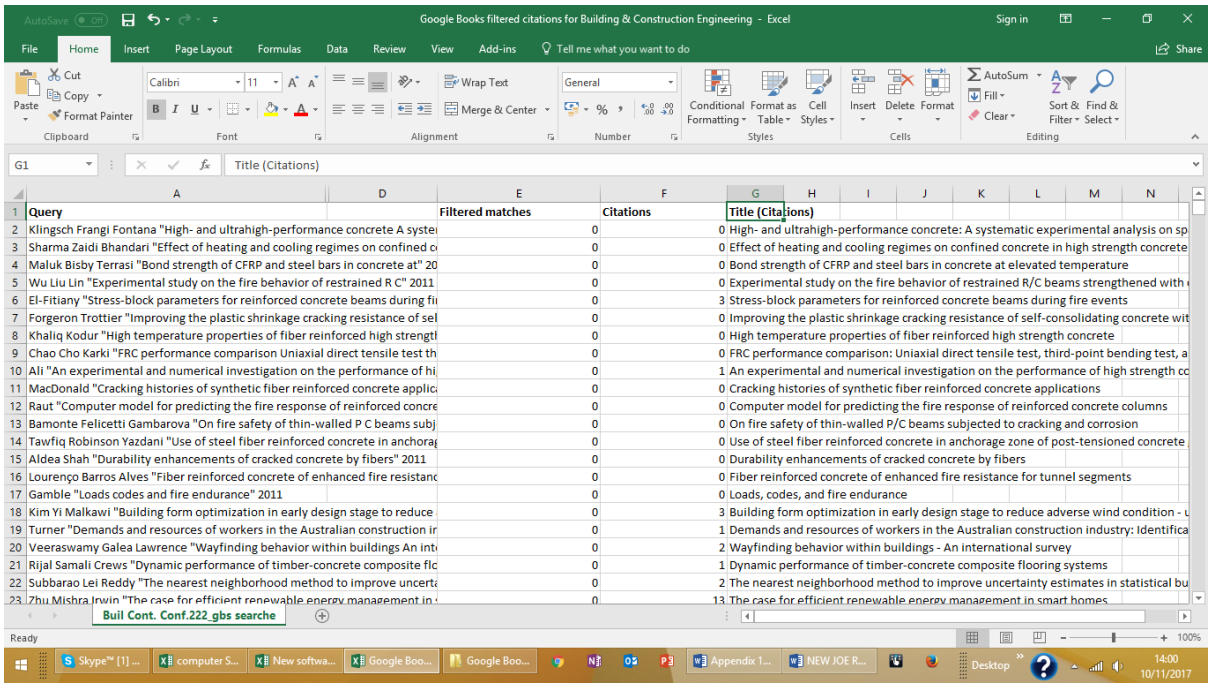
	Query	Filtered matches	Citations	Title
2	Gálvez-López "Real-time loop detection with bags of binai	3	6	Real-time loop detection with bags of binary words
3	Tadakuma Terada Ming "The mechanism of the linear load	0	0	The mechanism of the linear load-sensitive continuously variable transmission with the sph
4	Mukai Hirano Yoshida "Whole-body contact manipulation	2	18	Whole-body contact manipulation using tactile information for the nursing-care assistant rob
5	Rosa Herman Szewczyk "Laparoscopic optical biopsies In v	3	8	Laparoscopic optical biopsies: In vivo robotized mosaicing with probe-based confocal endom
6	Gowal Prorok Martinoli "Two-phase online calibration for	0	3	Two-phase online calibration for infrared-based inter-robot positioning modules
7	Williams "On adaptive underwater object detection" 2011	1	3	On adaptive underwater object detection
8	Carlési Michel Jouvencel "Generic architecture for multi-A	1	4	Generic architecture for multi-AUV cooperation based on a multi-agent reactive organizatior
9	Mathews Christensen O'Grady "Enhanced directional self-	1	3	Enhanced directional self-assembly based on active recruitment and guidance
10	Ghosh "The impact of different competence levels of care	2	4	The impact of different competence levels of care-receiving robot on children
11	Kapadia "Task-space control of extensible continuum man	1	12	Task-space control of extensible continuum manipulators
12	Tully Kantor Zenati "Shape estimation for image-guided su	0	13	Shape estimation for image-guided surgery with a highly articulated snake robot
13	Huang Kuan "Design of a new variable stiffness actuator ar	0	11	Design of a new variable stiffness actuator and application for assistive exercise control
14	Finio Galloway Wood "An ultra-high precision high bandw	1	8	An ultra-high precision, high bandwidth torque sensor for microrobotics applications
15	Chae Jeong Jo "Noninvasive brain-computer interface-bas	0	2	Noninvasive brain-computer interface-based control of humanoid navigation
16	Terekhov Mouret Grand "Stochastic optimization of a chai	0	2	Stochastic optimization of a chain sliding mode controller for the mobile robot maneuvering
17	Tahara Iwasa Naba "High-backdrivable parallel-link manip	0	2	High-backdrivable parallel-link manipulator with continuously variable transmission
18	Oki Abiko Nakanishi "Time-optimal detumbling maneuver	0	2	Time-optimal detumbling maneuver along an arbitrary arm motion during the capture of a ta
19	Biggs Ando Kotoku "A component supervisor for RT-Middl	0	0	A component supervisor for RT-Middleware using supervision trees
20	De Gea "Predictive compliance for interaction control of r	0	0	Predictive compliance for interaction control of robot manipulators
21	Pereira Santos Vassallo "A nonlinear controller for people	0	2	A nonlinear controller for people guidance based on omnidirectional vision
22	Bok Choi Jeong "Capturing city-level scenes with a synchro	0	5	Capturing city-level scenes with a synchronized camera-laser fusion sensor
23	Natrai Demonceaux Vasseur "Vision based attitude and al	0	7	Vision based attitude and altitude estimation for UAVs in dark environments

Appendix 13 Screenshot of Google Books filtered citation counts and Scopus citation counts for Software Engineering



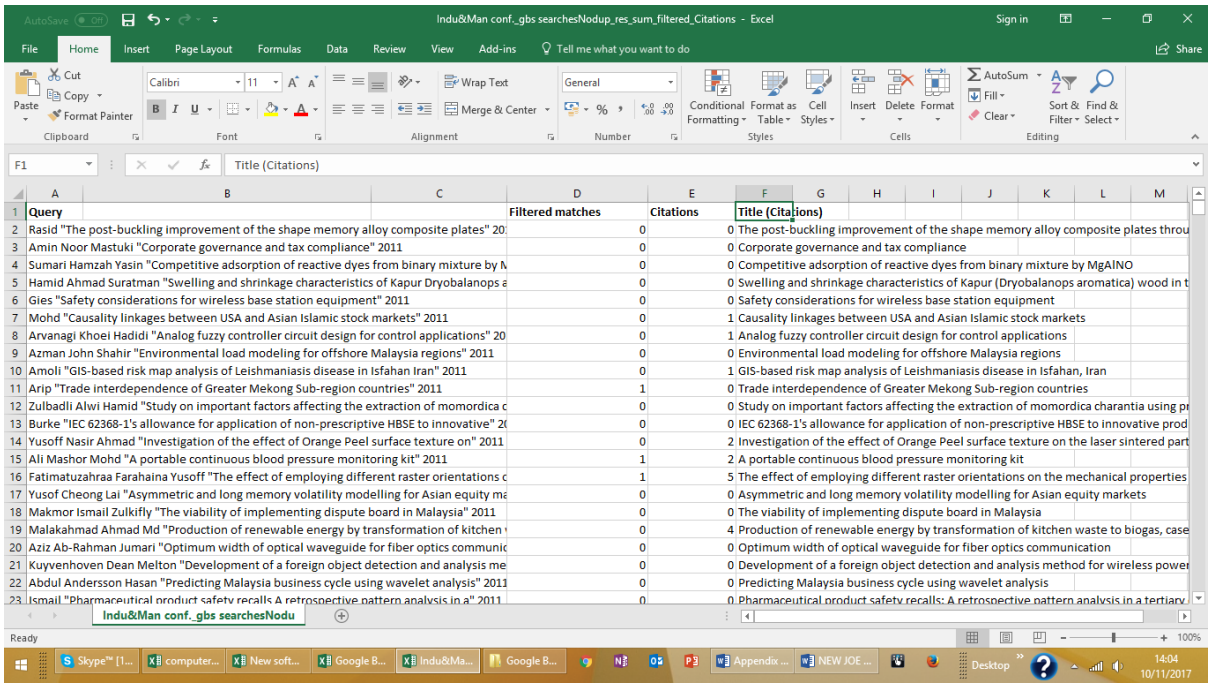
	Query	Filtered matches	Citations	Title (Citations)
2	Gálvez-López "Real-time loop detection with bags of binary words" 2011	3	6	Real-time loop detection with bags of binary words
3	Tadakuma Terada Ming "The mechanism of the linear load-sensitive continuously vari	0	0	The mechanism of the linear load-sensitive continuously variab
4	Mukai Hirano Yoshida "Whole-body contact manipulation using tactile information fo	2	18	Whole-body contact manipulation using tactile information for
5	Rosa Herman Szewczyk "Laparoscopic optical biopsies In vivo robotized mosaicing wit	3	8	Laparoscopic optical biopsies: In vivo robotized mosaicing with
6	Gowal Prorok Martinoli "Two-phase online calibration for infrared-based inter-robot	0	3	Two-phase online calibration for infrared-based inter-robot po
7	Williams "On adaptive underwater object detection" 2011	1	3	On adaptive underwater object detection
8	Carlési Michel Jouvencel "Generic architecture for multi-AUV cooperation based on a	1	4	Generic architecture for multi-AUV cooperation based on a mul
9	Mathews Christensen O'Grady "Enhanced directional self-assembly based on active re	1	3	Enhanced directional self-assembly based on active recruitmer
10	Ghosh "The impact of different competence levels of care-receiving robot on" 2011	2	4	The impact of different competence levels of care-receiving ro
11	Kapadia "Task-space control of extensible continuum manipulators" 2011	1	12	Task-space control of extensible continuum manipulators
12	Tully Kantor Zenati "Shape estimation for image-guided surgery with a highly articulat	0	13	Shape estimation for image-guided surgery with a highly articu
13	Huang Kuan "Design of a new variable stiffness actuator and application for" 2011	0	11	Design of a new variable stiffness actuator and application for a
14	Finio Galloway Wood "An ultra-high precision high bandwidth torque sensor for micro	1	8	An ultra-high precision, high bandwidth torque sensor for micro
15	Chae Jeong Jo "Noninvasive brain-computer interface-based control of humanoid nav	0	2	Noninvasive brain-computer interface-based control of human
16	Terekhov Mouret Grand "Stochastic optimization of a chain sliding mode controller fo	0	2	Stochastic optimization of a chain sliding mode controller for th
17	Tahara Iwasa Naba "High-backdrivable parallel-link manipulator with continuously var	0	2	High-backdrivable parallel-link manipulator with continuously v
18	Oki Abiko Nakanishi "Time-optimal detumbling maneuver along an arbitrary arm mot	0	2	Time-optimal detumbling maneuver along an arbitrary arm mot
19	Biggs Ando Kotoku "A component supervisor for RT-Middleware using supervision tre	0	0	A component supervisor for RT-Middleware using supervision t
20	De Gea "Predictive compliance for interaction control of robot manipulators" 2011	0	0	Predictive compliance for interaction control of robot manipula
21	Pereira Santos Vassallo "A nonlinear controller for people guidance based on omnidir	0	2	A nonlinear controller for people guidance based on omnidirec
22	Bok Choi Jeong "Capturing city-level scenes with a synchronized camera-laser fusion s	0	5	Capturing city-level scenes with a synchronized camera-laser fu
23	Natrai Demonceaux Vasseur "Vision based attitude and altitude estimation for UAVs i	0	7	Vision based attitude and altitude estimation for UAVs in dark

Appendix 14 Screenshot of Google Books filtered citation counts and Scopus citation counts for Building & Construction Engineering



Query	Filtered matches	Citations	Title (Citations)
Klingsch Frangi Fontana "High- and ultrahigh-performance concrete A syste	0	0	High- and ultrahigh-performance concrete: A systematic experimental analysis on sp
Sharma Zaidi Bhandari "Effect of heating and cooling regimes on confined c	0	0	Effect of heating and cooling regimes on confined concrete in high strength concrete
Maluk Bisby Terrasi "Bond strength of CFRP and steel bars in concrete at" 20	0	0	Bond strength of CFRP and steel bars in concrete at elevated temperature
Wu Liu Lin "Experimental study on the fire behavior of restrained R C" 2011	0	0	Experimental study on the fire behavior of restrained R/C beams strengthened with
El-Fitiany "Stress-block parameters for reinforced concrete beams during fi	0	3	Stress-block parameters for reinforced concrete beams during fire events
Forgeron Trottier "Improving the plastic shrinkage cracking resistance of sel	0	0	Improving the plastic shrinkage cracking resistance of self-consolidating concrete wit
Khaliq Kodur "High temperature properties of fiber reinforced high strengt	0	0	High temperature properties of fiber reinforced high strength concrete
Chao Cho Karki "FRC performance comparison Uniaxial direct tensile test th	0	0	FRC performance comparison: Uniaxial direct tensile test, third-point bending test, a
Ali "An experimental and numerical investigation on the performance of hi	0	1	An experimental and numerical investigation on the performance of high strength cc
MacDonald "Cracking histories of synthetic fiber reinforced concrete applic	0	0	Cracking histories of synthetic fiber reinforced concrete applications
Raut "Computer model for predicting the fire response of reinforced concre	0	0	Computer model for predicting the fire response of reinforced concrete columns
Bamonte Felicetti Gambarova "On fire safety of thin-walled P C beams subj	0	0	On fire safety of thin-walled P/C beams subjected to cracking and corrosion
Tawfiq Robinson Yazdani "Use of steel fiber reinforced concrete in anchorag	0	0	Use of steel fiber reinforced concrete in anchorage zone of post-tensioned concrete
Aldea Shah "Durability enhancements of cracked concrete by fibers" 2011	0	0	Durability enhancements of cracked concrete by fibers
Lourenço Barros Alves "Fiber reinforced concrete of enhanced fire resistanc	0	0	Fiber reinforced concrete of enhanced fire resistance for tunnel segments
Gamble "Loads codes and fire endurance" 2011	0	0	Loads, codes, and fire endurance
Kim Yi Malkawi "Building form optimization in early design stage to reduce	0	3	Building form optimization in early design stage to reduce adverse wind condition -
Turner "Demands and resources of workers in the Australian construction ir	0	1	Demands and resources of workers in the Australian construction industry: Identifica
Veeraswamy Galea Lawrence "Wayfinding behavior within buildings An int	0	2	Wayfinding behavior within buildings - An international survey
Rijal Samali Crews "Dynamic performance of timber-concrete composite flo	0	1	Dynamic performance of timber-concrete composite flooring systems
Subbarao Lei Reddy "The nearest neighborhood method to improve uncerta	0	2	The nearest neighborhood method to improve uncertainty estimates in statistical bu
Zhu Mishra Irwin "The case for efficient renewable energy management in	0	13	The case for efficient renewable energy management in smart homes

Appendix 15 Screenshot of Google Books filtered citation counts and Scopus citation counts for Industrial & Manufacturing Engineering



Query	Filtered matches	Citations	Title (Citations)
Rasid "The post-buckling improvement of the shape memory alloy composite plates" 20	0	0	The post-buckling improvement of the shape memory alloy composite plates throu
Amin Noor Mastuki "Corporate governance and tax compliance" 2011	0	0	Corporate governance and tax compliance
Sumari Hamzah Yasin "Competitive adsorption of reactive dyes from binary mixture by M	0	0	Competitive adsorption of reactive dyes from binary mixture by MgAlNO
Hamid Ahmad Suratman "Swelling and shrinkage characteristics of Kapur Dryobalanops a	0	0	Swelling and shrinkage characteristics of Kapur (Dryobalanops aromatica) wood in t
Gies "Safety considerations for wireless base station equipment" 2011	0	0	Safety considerations for wireless base station equipment
Mohd "Causality linkages between USA and Asian Islamic stock markets" 2011	0	1	Causality linkages between USA and Asian Islamic stock markets
Arvanagi Khoei Hadidi "Analog fuzzy controller circuit design for control applications" 20	0	1	Analog fuzzy controller circuit design for control applications
Azman John Shahir "Environmental load modeling for offshore Malaysia regions" 2011	0	0	Environmental load modeling for offshore Malaysia regions
Amoli "GIS-based risk map analysis of Leishmaniasis disease in Isfahan Iran" 2011	0	1	GIS-based risk map analysis of Leishmaniasis disease in Isfahan, Iran
Arip "Trade interdependence of Greater Mekong Sub-region countries" 2011	1	0	Trade interdependence of Greater Mekong Sub-region countries
Zulbadli Alwi Hamid "Study on important factors affecting the extraction of momordica c	0	0	Study on important factors affecting the extraction of momordica charantia using pr
Burke "IEC 62368-1's allowance for application of non-prescriptive HBSE to innovative" 2(	0	0	IEC 62368-1's allowance for application of non-prescriptive HBSE to innovative prod
Yusoff Nasir Ahmad "Investigation of the effect of Orange Peel surface texture on" 2011	0	2	Investigation of the effect of Orange Peel surface texture on the laser sintered part
Ali Mashor Mohd "A portable continuous blood pressure monitoring kit" 2011	1	2	A portable continuous blood pressure monitoring kit
Fatimatuzahraa Farahaina Yusoff "The effect of employing different raster orientations c	1	5	The effect of employing different raster orientations on the mechanical properties
Yusof Cheong Lai "Asymmetric and long memory volatility modelling for Asian equity ma	0	0	Asymmetric and long memory volatility modelling for Asian equity markets
Makmor Ismail Zulkifly "The viability of implementing dispute board in Malaysia" 2011	0	0	The viability of implementing dispute board in Malaysia
Malakahmad Ahmad Md "Production of renewable energy by transformation of kitchen	0	4	Production of renewable energy by transformation of kitchen waste to biogas, case
Aziz Ab-Rahman Jumari "Optimum width of optical waveguide for fiber optics communic	0	0	Optimum width of optical waveguide for fiber optics communication
Kuyvenhoven Dean Melton "Development of a foreign object detection and analysis me	0	0	Development of a foreign object detection and analysis method for wireless power
Abdul Andersson Hasan "Predicting Malaysia business cycle using wavelet analysis" 2011	0	0	Predicting Malaysia business cycle using wavelet analysis
Ismail "Pharmaceutical product safety recalls A retrospective pattern analysis in a" 2011	0	0	Pharmaceutical product safety recalls: A retrospective pattern analysis in a tertiary

Appendices 16 to 22 show how the 'R software was used to calculate confidence intervals for both journal articles and conference papers for Mendeley readers, Google Patents citations, Wikipedia citations and citations to Google books for all the selected subject categories (Computer Science applications, Computer software Engineering, Building & construction engineering and Industrial and Manufacturing Engineering.

The confidence interval is a way to show what uncertainty is with a certain statistic either from a poll or survey. For example, a poll might state that there are is a 95% confidence interval of 0.546 and 0.560. That means if the poll is repeated using the same techniques, 95% of the time the population parameter will fall within the interval estimates (0.546 and 0.560) 95% of the time.

Appendix 16 Confidence Interval for Mendeley readers (Journal articles)

```
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)
Natural language support but running in an English locale
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.560,10000)
[1] 0.546397
> ffu(0.560,10000)
[1] 0.5733077
> # Confidence interval for Computer Science Applications
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.572,10000)
[1] 0.558663
> ffu(0.572,10000)
[1] 0.5850413
> #Confidence interval for Computer Software Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.662,8433)
[1] 0.649838
> ffu(0.662,8433)
[1] 0.6738231
> #Confidence interval for Building & Construction Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.660,10000)
[1] 0.6709212
> ffl(0.660,10000)
[1] 0.6487925
```

> # Confidence interval for Industrial & Manufacturing Engineering

Appendix 17 Confidence Interval for Mendeley readers (Conference papers)

R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)
Natural language support but running in an English locale
> ffu<-function(t,n) tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n) tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(.439,9999)
[1] 0.4546888
> ffl(0.439,9999)
[1] 0.4230388
> #Confidence interval for Computer science applications
> ffu<-function(t,n) tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n) tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(.437,9974)
[1] 0.452743
> ffl(0.437,9974)
[1] 0.4209846
> #Confidence interval for Computer Software Engineering
> ffu<-function(t,n) tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n) tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.143,4750)
[1] 0.1150278
> ffu(.143,4750)
[1] 0.1707456
> #Confidence interval for Building & Construction Engineering
> ffu<-function(t,n) tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n) tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.168,9999)
[1] 0.1488889
> ffu(.168,9999)
[1] 0.1869857
> #Confidence interval for Industrial & Manufacturing Engineering

Appendix 18 Confidence interval for Google Patents (Conference papers)

R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
Natural language support but running in an English locale
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.072,6698)

[1] 0.0957844
> ffl(0.072,6698)
[1] 0.04813343
> # confidence interval for computer Science Application
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.018,8083)
[1] 0.03978565
> ffl(0.018,8083)
[1] -0.003802751
> # Confidence interval for Software Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(-0.015,5651)
[1] 0.01107849
> ffl(-0.015,5651)
[1] -0.0410581
> # Confidence interval for Industrial & Manufacturing Engineering

Appendix 19 Confidence interval for Google Patents (Journal articles)

> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffu<(0.005,8407)
[1] 0.02637419
> ffl(0.005,8407)
[1] -0.01637876
> # Confidence interval for Building & Construction Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(-0.006,7355)
[1] 0.01685713
> ffl(-0.006,7355)
[1] -0.02885086
> # Confidence interval for Industrial& Manufacturing Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.2,8150)
[1] 0.2207529
> ffl(0.2,8150)
[1] 0.1790661
> # Confidence interval for Computer Science Applications
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.015,8232)

[1] 0.03659119
> ffl(0.015,8232)
[1] -0.006605185
> #Confidence interval for Software Engineering

Appendix 20 Confidence interval for Wikipedia citations (Conference papers)

R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
Natural language support but running in an English locale
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.274,6700)
[1] 0.2517055
> ffu(0.274,6700)
[1] 0.2960039
> #Confidence interval for computer Science Applications
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(-0.0010,8085)
[1] -0.02279807
> ffu(-0.0010,8085)
[1] 0.02079902
> #Confidence interval for Software Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(-0.017,5650)
[1] -0.04305738
> ffu(-0.017,5650)
[1] 0.00908049
> #Confidence interval for Industrial& Manufacturing Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.040,1753)
[1] -0.006831501
> ffu(0.040,1753)
[1] 0.08665642
> #Confidence interval for Industrial& Manufacturing Engineering

Appendix 21 Confidence interval for Wikipedia citations (Journal articles)

> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.056,5912)
[1] 0.03055155

```
> ffu(0.056,5912)
[1] 0.08137589
> #confidence interval for Computer Science Applications
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.020,8231)
[1] -0.001605051
> ffu(0.020,8231)
[1] 0.04158639
> #confidence interval for Computer Software Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.012,7354)
[1] -0.01085935
> ffu(0.012,7354)
[1] 0.03484681
> #confidence interval for Industrial& Manufacturing Engineering
> ffu<-function(t,n)tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n)tanh(atanh(t)-1.96/sqrt(n-3))
> ffl(0.002,8406)
[1] -0.01937911
> ffu(0.002,8406)
[1] 0.02337728
> #Confidence interval for Building & Construction Engineering
```

Appendix 22 Confidence interval for Google Books (Conference papers)

```
R version 3.2.0 (2015-04-16) -- "Full of Ingredients"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
 Natural language support but running in an English locale
> ffu<-function(t,n) tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n) tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.305,9982)
[1] 0.3226873
> ffl(0.305,9982)
[1] 0.2870998
> #confidence interval for Computer Science Applications
> ffu<-function(t,n) tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n) tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.285,9990)
[1] 0.3029173
> ffl(0.285,9990)
[1] 0.2668813
> #confidence interval for Software Engineering
```

```
> ffu<-function(t,n) tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n) tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.092,9990)
[1] 0.1114092
> ffl(0.092,9990)
[1] 0.0725206
> #Confidence interval for Industrial & Manufacturing Engineering
> ffu<-function(t,n) tanh(atanh(t)+1.96/sqrt(n-3))
> ffl<-function(t,n) tanh(atanh(t)-1.96/sqrt(n-3))
> ffu(0.000,4602)
[1] 0.02889372
> ffl(0.000,4602)
[1] -0.02889372
> #Confidence Interval for Building & Construction Engineering
```